1999

Discrimination as Accident

Amy L. Wax
University of Pennsylvania Carey Law School

Discrimination as Accident

AMY L. WAX*

TABLE OF CONTENTS

---

* Associate Professor of Law, University of Virginia Law School. I wish to thank Ken Abraham, Kate Bartlett, Kingsley Browne, Jacques deLisle, John Donohue, Cynthia Estlund, Charles Goetz, Linda Krieger, Stephen J. Morse, John Monahan, Glen Robinson, Rip Verkerke, Michael Selmi, Bill Stuntz, Timothy Wilson, and participants in the Legal Studies Workshop at the University of Virginia Law School, the lunchtime workshop series at the University of Virginia Department of Psychology, and the 1999 annual meeting of the Law and Economics Association, for helpful comments and criticism. I am grateful for financial support from the University of Virginia Program for Employment and Labor Law Studies. John Adair and Marcelle Morel provided excellent research assistance.

INTRODUCTION

Scholars addressing the problem of discrimination against socially disfavored groups have distinguished between two types of bias in a variety of social settings: "conscious," deliberate, or purposeful animus, and "unconscious," inadvertent, or automatic forms of bias.[1] Although the idea that some group-based bias may be "unconscious" or unintentional has been around for some time, the subject has received growing attention. Some commentators have gone so far as to suggest that, as overt bigotry has waned in response to antidiscrimination laws and evolving social mores, unintentional or "unconscious" discrimination has become the most pervasive and important form of bias operating in society today.[2] This is especially

---

1. *See, e.g.*, JODY DAVID ARMOUR, NEGROPHOBIA AND REASONABLE RACISM: THE HIDDEN COSTS OF BEING BLACK IN AMERICA 68-80 (1997); Paul Brest, *The Supreme Court, 1975 Term-Forward: In Defense of the Antidiscrimination Principle*, 90 HARV. L. REV. 1, 6-7 (1976); Judith Olans Brown et al., *Some Thoughts About Social Perception and Employment Discrimination Law: A Modest Proposal for Reopening the Judicial Dialogue*, 46 EMORY L.J. 1487, 1493-97 (1997); Martha Chamallas, *The Architecture of Bias: Deep Structures in Tort Law*, 146 U. PA. L. REV. 463, 466-67 (1998); Peggy C. Davis, *Law as Microaggression*, 98 YALE L.J. 1559, 1560 (1989); Barbara J. Flagg, *"Was Blind, But Now I See": White Race Consciousness and the Requirement of Discriminatory Intent*, 91 MICH. L. REV. 953 *passim* (1993); Sheri Lynn Johnson, *Unconscious Racism and the Criminal Law*, 73 CORNELL L. REV. 1016 *passim* (1988); Linda Hamilton Krieger, *Civil Rights Perestroika: Intergroup Relations After Affirmative Action*, 86 CAL. L. REV. 1251, 1279, 1286-91 (1998) [hereinafter Krieger I]; Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1164 (1995) [hereinafter Krieger II]; Charles R. Lawrence III, *The Id, The Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317 *passim* (1987); Anne C. McGinley, *Rethinking Civil Rights and Employment at Will: Toward a Coherent Discharge Policy*, 57 OHIO ST. L.J. 1443, 1463-73 (1996); David Benjamin Oppenheimer, *Negligent Discrimination*, 141 U. PA. L. REV. 899, 900-17 (1993); Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. REV. 1251, 1283-89 (1995); David A. Strauss, *Discriminatory Intent and the Taming of* Brown, 56 U. CHI. L. REV. 935, 960-62 (1989); Jessie Allen, Note, *A Possible Remedy for Unthinking Discrimination*, 61 BROOK. L. REV. 1299, 1311-15 (1995); Pamela S. Karlan, Note, *Discriminatory Purpose and Mens Rea: The Tortured Argument of Invidious Intent*, 93 YALE L.J. 111, 124-28 (1983).

2. *See* ARMOUR, *supra* note 1, at 72-77 (stressing the importance of unconscious bias in multiple spheres); VIRGINIA VALIAN, WHY SO SLOW? THE ADVANCEMENT OF WOMEN 1-9 (1998) (suggesting that the accumulation of small disadvantages and setbacks from unconscious bias is the dominant factor impeding women's workplace advancement); Brown et al., *supra* note 1, at 1492-1503 (same); Krieger I, *supra* note 1, at 1279, 1286-91 (same); Krieger II, *supra* note 1, at 1164-69 (arguing for the central importance of "subtle unconscious bias" in the workplace); *see also, e.g.*, David K. Shipler, *Seeing Through Camouflaged Racism*, WASH. POST, Oct. 15, 1997,

true, it is claimed, in the workplace. It is not just that employers are more careful about engaging in overtly discriminatory behavior, or that they guard against open declarations of their prejudiced sentiments.[3] Rather, the bulk of workplace discrimination has taken on an entirely new form.[4]

The focus of this Article is on a type of discriminatory conduct in the workplace that will be termed "unconscious disparate treatment." Under existing laws governing discrimination on the job, disparate treatment results when an employer disfavors a worker "because of" or "based on" that person's membership in a protected group.[5] An employer or his agent may know full well that he is treating an employee less well because of that employee's race or sex. In that case, the difference in treatment is deliberate or "conscious." But the employer may be unaware that he is treating the employee differently than others and oblivious to the basis for that difference. Through the operation of cognitive mechanisms that the decisionmaker can neither observe nor control, his perception of a worker's race or sex may distort the application of neutral and reasonable criteria used to evaluate that employee. Although the decisionmaker might think he is being "fair," the inadvertent bias might alter the outcome of a decision about the employee. For example, a supervisor might unconsciously place more weight on errors in grammar or spelling in a memo prepared by a Hispanic clerk than in a document submitted by an Anglo counterpart. Or he might view a female employee's restraint in a client meeting as manifesting a lack of aggressiveness rather than prudence or good judgment. In both cases, the "difference" in the employer's reaction ultimately depends upon the employer's identification of the employee as belonging to a particular group. This conduct fits the framework of disparate treatment because observations regarding the worker's protected trait are causally linked to less favorable judgments, which may result in adverse treatment in the workplace.

This Article seeks to address how the law should respond to the problem of unconscious disparate treatment in the workplace. It is concerned primarily with the problem of inadvertent bias in employee performance appraisals. That form of bias can result from reflexive or unthinking distortions in the application of neutral and seemingly reasonable criteria to the assessment of employees from disfavored groups. The analysis is concerned primarily with inadvertent bias against minorities and women in the workplace.[6] Specifically, this Article asks whether unconscious

---

at 21.

3. *See* JOHN J. DONOHUE III, FOUNDATIONS OF EMPLOYMENT DISCRIMINATION LAW 260 (1997) (noting the decline in "direct and indisputable evidence of labor market discrimination," and suggesting that the disappearance of open proclamations such as "No Irish Need Apply" has been "far more thorough[] than [the elimination] of discriminatory conduct on the part of employers").

4. *See, e.g.*, Krieger II, *supra* note 1, at 1164; Selmi, *supra* note 1, at 1283.

5. The source of the proscription against disparate treatment lies in the language of Title VII, which forbids discrimination "because of" traits such as race or sex. *See* Civil Rights Act of 1964 § 703(a), 42 U.S.C. § 2000e-2 (1994); *see also* Civil Rights Act of 1866, ch. 31, § 1, 14 Stat. 27 (1866) (codified as amended at 42 U.S.C. §§ 1981, 1983 (1994)).

6. Although this Article deals with a type of inadvertent bias that gives rise to *disparate treatment*, its scope is not meant to suggest that there are no other types of workplace conduct that might properly be categorized as "unconscious discrimination." There may be other important patterns or conditions, not qualifying as disparate treatment, that make it harder for women and

disparate treatment should be addressed within the framework established by current laws covering workplace discrimination, such as Title VII, which allow individuals and groups to sue employers for equitable and compensatory relief.[7] The method of the Article is economic in its concern with rational actors' responses to the incentives created by the law's allocation of costs and benefits among parties to social interactions and with whether those responses "will promote, or fail to promote, social welfare."[8]

This Article proceeds from the assumption that discrimination against persons based on group identity inflicts a form of harm on victims. Title VII and other provisions designed to combat workplace discrimination create a system of liability, administered by courts, to remedy the harms individual workers suffer from discriminatory treatment. Title VII represents a judgment that society should take steps both to eliminate this harm and to provide equitable and legal compensation to victims.

Although some scholars have found it useful to view statutes combating workplace discrimination as creating liability systems designed to remedy individual injury, harm, or loss,[9] the problem of unconscious discrimination has not been systematically analyzed in these terms. Likewise, scholars have not seen fit to view inadvertent discrimination as a form of workplace "accident." That classification is apt, because unconscious discrimination is an inadvertent and unpredictable event that inflicts an undesirable and costly injury on employees as the side effect of the otherwise useful activity of employing a workforce in a productive enterprise. Moreover, the analogy is fruitful in inviting the use of familiar concepts of accident law to shed light on the consequences of legal interventions that ostensibly are designed to shift the costs of harms to employers from employees.

Using familiar principles of accident law, this Article concludes that extending the framework created by existing antidiscrimination statutes to cover unconscious workplace disparate treatment is not a good idea because it is unlikely to serve the principal goals of a liability scheme—deterrence, compensation, insurance—in a

---

minorities to get ahead. Many such practices are best classified as problems of disparate impact, although some would resist that categorization as well. As discussed below, arguments for de-emphasizing liability based on categories of discrimination established under Title VII include the difficulty of categorizing and distinguishing different sources of obstacles women and minorities face, as well as the wisdom of developing a unified approach to those problems that does not fit easily within established remedial forms. *See infra* Part II.B (discussing complex structural obstacles to outgroup advancement as an uneasy fit with established concepts of "discrimination"); *see also* sources cited *supra* note 1.

7. This Article does not undertake an analysis of statutes imposing liability for discrimination based on age or disability, but rather focuses on race, sex, and ethnicity. But much of the analysis here might be usefully applied in those contexts.

8. A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 873 (1998).

9. *See, e.g.*, Mark S. Brodin, *The Role of Fault and Motive in Defining Discrimination: The Seniority Question Under Title VII*, 62 N.C. L. REV. 943, 978-85 (1984); J. Hoult Verkerke, *Notice Liability in Employment Discrimination Law*, 81 VA. L. REV. 273, 288 (1995); Mark C. Weber, *Beyond* Price Waterhouse v. Hopkins: *A New Approach to Mixed Motive Discrimination*, 68 N.C. L. REV. 495, 496 (1990).

cost effective manner. Holding employers liable will not deter the harm of unconscious disparate treatment unless employers and their agents can find ways to reduce that harm. The nature of the phenomenon of unconscious bias is such that there are no known steps that employers can reliably take to control biases that may distort the kinds of discretionary or subjective social judgments that must inevitably be made in the course of managing personnel in the modern workplace. For this reason, little or no cost-justified deterrence of unconscious bias can be expected to result from imposing liability on employers for this type of conduct.

The Article goes on to suggest that the intransigence of unconscious bias will lead employers to respond to the threat of liability for inadvertent bias by overinvesting in measures that may reduce their exposure to liability, but will do little to purge unconscious bias from the decisionmaking process. As noted, employers have little effective control over unconscious bias. They do, however, have comprehensive control over both actionable (i.e., discriminatory) and nonactionable (nondiscriminatory) causes of adverse decisions taken against their employees (such as refusal to hire, demotion, discipline, and termination). They will therefore be tempted to respond to the threat of liability by, in effect, reducing the number of adverse actions taken against protected employees in a manner that need not require the precise elimination of inadvertent bias from the evaluation process. Alternatively, employers may take steps to override some of the potential *effects* of such bias—such as avoiding adverse decisions against protected groups—but in ways that may bear little relationship to the overall incidence of the targeted harm of inadvertently biased decisionmaking, and may actually increase the incidence of discrimination by introducing fresh forms of bias. Firms may, for example, adopt various types of diversity awareness, diversity action, or affirmative action programs. Employers will adopt these programs, however, only if they reduce the risk of liability. Moreover, if introducing these modifications has the effect of reducing employers' expected liability by an amount that exceeds their cost, employers will take these steps regardless of whether bias in the workplace has actually been reduced in a cost-effective manner or at all. This response will be inefficient because the amount expended on "precautions" against liability will fall short of abating the targeted harm to a degree that justifies such expenditures.

There are two reasons why any reduction in liability for unconscious bias that employers can expect from the above-mentioned "precautions" may well exceed actual harm reduction: First, there are no known methods for effectively controlling unconscious bias in the workplace. Therefore, any steps that employers take are as likely to be ineffective as effective. Second, unconscious bias is an intermittent, subtle, and elusive phenomenon. For this reason, fact-finders will be prone to error in assessing the actual incidence of unconscious bias, and employers may well succeed in using the above-mentioned workplace reforms to refute claims of bias or to persuade fact-finders that unconscious bias has abated. Moreover, even setting aside the question of whether such workplace modifications are likely to be an efficient response to holding employers liable for inadvertent bias, a cumbersome and proof-intensive liability system is a particularly wasteful way to force the adoption of these types of measures.

Any argument in favor of making the employer pay for unconscious bias also requires assuming that, as between the firm and its employees, the former can more cheaply reduce the amount of unconscious disparate treatment in the workplace than

the latter. This Article suggests that this assumption is unwarranted: there are reasons to believe that employees are in a better position to minimize the incidence of biased assessments, in which case liability principles would argue for leaving the costs of such bias where they fall.

Finally, attaching liability to unconscious forms of discriminatory treatment will do little to advance the cause of fair and accurate compensation for victims. If the goal is corrective or compensatory justice for victimized individuals, a liability system for unconscious disparate treatment will be a failure. Legal commentators have spoken of unconscious bias as "subtle" or elusive, but without giving precise content to these terms.[10] Research in social and cognitive psychology permits a more specific, albeit speculative, formulation: unconscious bias appears to occur unpredictably and sporadically in social interactions, and only sometimes appears to make an important difference to their outcomes. Such biases, if they operate at all, may therefore affect the workplace "bottom line" only erratically or infrequently. This Article argues that if unconscious bias is indeed "subtle" in these respects, determinations of liability will very often be in error. Depending on the placement of burdens and the chosen standard of proof, employers and firms will either be dramatically undercharged or overcharged for their misconduct, and compensation will often be paid to the wrong employees, with deserving victims receiving nothing or too little and the uninjured receiving too much. These errors in compensation will generally be greater in magnitude and frequency than in cases involving conscious bias. Although a probabilistic approach to liability would appear to offer some hope for improving on this situation, that promise is illusory. A probabilistic system will justify itself neither in producing well-calibrated risk reduction nor in directing compensation to the right persons. Because patterns of compensation to individuals under any of these schemes will inevitably be haphazard and bear little relation to actual victimization, those patterns will come to mimic the effects of affirmative action in the allocation of rewards and deprivation. But once again, a judicially administered system of individualized liability, which expends elaborate resources on seeking to identify and verify an elusive individualized form of harm, is a wasteful and cumbersome way to implement affirmative action in the workplace. That goal can be accomplished more efficiently by other means.

This Article concludes that aggressively attacking unconscious workplace bias with a scheme modeled on Title VII is unlikely to serve the traditional purposes of a liability system very well. Because this approach will prove both burdensome and expensive, this Article suggests that the law should attack unconscious bias in the workplace in other ways, or leave it to the play of extralegal forces.

---

10. *See* sources cited *supra* notes 1-2.

I. THE PROBLEM OF UNCONSCIOUS BIAS IN THE WORKPLACE

A. Unconscious Bias as "Mental Contamination"

A law professor has two students in his class. One is very beautiful and the other quite plain. The professor assigns a grade based on an in-class performance, a paper, and class participation. He does not plan to assign grades based on students' physical appearance. Nevertheless, he gives the first student a higher grade than she would have received if she were as plain as the second. The second student receives a lower grade than she would have obtained had she been as pretty as the first.

A manager is asked to supervise a black employee. The supervisor observes the employee's performance in the workplace over many months, taking in a range of information about him and the work he produces. The supervisor is asked to fill out an evaluation form for the employee, describing and ranking his on-the-job performance and the quality of his work based on a range of criteria. Although the supervisor believes that he is ignoring the employee's race, he is not: in fact, he gives the black employee a less favorable evaluation than the employee would have received had he been white.

In both cases, one possible explanation for the scenario is that the decisionmaker (supervisor or professor) is influenced in his subjective assessment by a "stimulus" or observed attribute—that is, physical attractiveness or race—of the person being evaluated. The evaluator is unaware of this influence and, indeed, may fervently wish to avoid taking the "stimulus" into account. Psychologists Timothy Wilson and Nancy Brekke have described this type of situation as a form of "mental contamination," which they define as "an unwanted judgment, emotion, or behavior" influenced by "mental processing that is unconscious or uncontrollable."[11] The influence is unwanted if the person making the judgment would not wish a particular factor to play a role in his judgment.[12] One example of "mental contamination" is the "halo effect"—the tendency to give beautiful people a higher rating or to evaluate their performance more favorably, even if the overt criteria for appraising the performance have nothing to do with physical appearance.[13]

---

11. *See* Timothy D. Wilson & Nancy Brekke, *Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations*, 116 PSYCHOL. BULL. 117, 117 (1994); *see also* John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effect*, *in* DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY 361, 363 (Shelly Chaiken & Yaacov Trope eds., 1999) (describing how "[t]he mere perception of easily discernible group features" might "influence judgments of a group member in an unintended fashion, outside of a perceiver's awareness").

12. *See id.*; *see also* Krieger I, *supra* note 1, at 1285-91 (discussing and endorsing Wilson and Brekke's "mental contamination" paradigm).

13. *See* Wilson & Brekke, *supra* note 11, at 117; *see also* Note, *Facial Discrimination: Extending Handicap Law to Employment Discrimination on the Basis of Physical Appearance*, 100 HARV. L. REV. 2035, 2037-42 (1987) (describing evidence of discrimination based on physical appeal).

Mental contamination is but one form of cognitive bias that has been observed by experimental social psychologists under controlled or "laboratory" conditions.[14] Researchers have speculated that some types of mental contamination may result from the process of unconscious stereotyping. Stereotypes are abstract structures of knowledge or understandings that link group membership to a set of traits or behavioral characteristics. Once stereotypes are activated, "they tend to influence the construal of information about the target. As a result, the mental representation that is formed of the target person is likely to be influenced by stereotypical assumptions, even if the available information provides no direct substantiation for

14. Contamination biases have been observed in three basic types of experiments. In the first type, experimental subjects are asked to evaluate a person or a person's performance. The rating is to be based on various forms of information or experience, including an interview, a paper summary (i.e., a résumé), an interaction with the person, or an opportunity to observe the person in action (e.g., participating in a group discussion). Some experiments are specifically designed to investigate the possibility of inadvertent bias in workplace evaluations. The typical experimental design attempts to hold the performance constant, but varies a key attribute of the person (race, sex, age). Oddly enough, studies of this type that look at sex bias far outnumber those that explore responses to race. *See, e.g.*, Richard F. Martell, *Sex Bias at Work: The Effects of Attentional and Memory Demands on Performance Ratings of Men and Women*, 21 J. APPLIED SOC. PSYCHOL. 1939, 1940 (1991); John B. McConahay, *Modern Racism and Modern Discrimination: The Effects of Race, Racial Attitudes, and Context on Simulated Hiring Decisions*, 9 PERSONALITY SOC. PSYCHOL. BULL. 551, 555-58 (1983); Veronica F. Nieva & Barbara A. Gutek, *Sex Effects on Evaluation*, 5 ACAD. MGMT. REV. 267, 267 (1980) (describing a "common research paradigm used in studies of evaluation bias" as involving "the description of hypothetical persons who are identical except for sex, on whom evaluation judgments and personnel decisions are made"); Janet Swim et al., *Joan McKay Versus John McKay: Do Gender Stereotypes Bias Evaluations?*, 105 PSYCHOL. BULL. 409, 419-24 (1989); Henry L. Tosi & Steven W. Einbender, *The Effects of the Type and Amount of Information in Sex Discrimination Research: A Meta-Analysis*, 28 ACAD. OF MGMT. J. 712, 713-19 (1985) (describing analysis of studies).

In the second type of experiment (known as a "priming" experiment), an experimental subject's response is assessed following his exposure to an idea or influence. For example, it has been observed that people's decisions about whether to act cooperatively or competitively towards others depends on prior exposure to news broadcasts about pro or antisocial acts. *See* Wilson & Brekke, *supra* note 11, at 117. Also, experimental subjects' accounts of the meaning of videotapes of ambiguous encounters between blacks and whites are influenced by whether those subjects have been subliminally exposed to words that reflect stereotypes about black Americans. *See id.*; *see also* Irene V. Blair & Mahzarin R. Banaji, *Automatic and Controlled Processes in Stereotype Priming*, 70 J. PERSONALITY & SOC. PSYCHOL. 1142, 1143-58 (1996) (describing priming-type experiments).

Finally, psychologists have observed that people's responses to a social situation—such as a shopper dropping a bag of groceries on the sidewalk—will sometimes vary by the race or sex of the person encountered. *See, e.g.*, Faye Crosby et al., *Recent Unobtrusive Studies of Black and White Discrimination and Prejudice: A Literature Review*, 87 PSYCHOL. BULL. 546, 549 (1980).

these assumptions."[15] If the process of stereotyping is unconscious, an individual will be unaware that stereotypical expectancies are at work in his social judgments.

The operation of unconscious mechanisms stands in contrast to discriminatory treatment that is conscious or deliberate. In that case, the actor means to alter his judgment of a person precisely because that person possesses a particular attribute or trait. A person who discriminates consciously is aware of what he is doing: he knows that the factor of race is "making a difference" in his evaluations of others. The reasons for his actions are transparent to him, which gives rise to the capacity for self-report through introspection.[16] To paraphrase Justice Brennan, a person who is consciously discriminating, if asked about why he took a particular decision, is able to answer truthfully "because of race."[17]

This discussion of unconscious stereotyping suggests how employment relations might create opportunities for superiors to discriminate unconsciously against employees. Supervisors and employers purport to evaluate employees according to facially neutral criteria and often strive to apply those criteria in an evenhanded way. But if they have knowledge of the race or sex of the person being evaluated (which they ordinarily do), their judgments could possibly be affected by cognitive biases that are triggered by that knowledge. That would happen if the employer or his agent inadvertently applied categorical mental assumptions about blacks or women in a way that colors the evaluation of affected employees, leading to distortions in judgment and a less favorable evaluation. But the employer will not realize that these cognitive mechanisms are at work and will be oblivious to the way in which the application of neutral performance criteria, which he is attempting to apply in good faith, is skewed by his unconscious stereotypes. Such routine distortions of seemingly benign appraisals could potentially occur at all stages of the employment relationship, affecting decisions whether to hire, promote, discipline, assign responsibility, allocate rewards and benefits, or terminate the relationship altogether. The potential for these types of cognitive mechanisms to play a role would be greatest when assessments have an important subjective component—and especially where employers are making complex, multifactorial,

---

15. Galen V. Bodenhausen & C. Neil Macrae, *Stereotype Activation and Inhibition*, *in* 11 ADVANCES IN SOCIAL COGNITION: STEREOTYPE ACTIVATION AND INHIBITION 1, 16 (Robert S. Wyer, Jr. ed., 1998) [hereinafter ADVANCES IN SOCIAL COGNITION]; *see* David L. Hamilton & Jeffrey W. Sherman, *Stereotypes*, *in* 2 HANDBOOK OF SOCIAL COGNITION 1, 3 (Robert S. Wyer, Jr. & Thomas K. Srull eds., 2d ed. 1994) (Stereotypes "act as expectancies that guide the processing of information about the group as a whole and about group members."); David L. Hamilton & Tina K. Trolier, *Stereotypes and Stereotyping: An Overview of the Cognitive Approach*, *in* PREJUDICE, DISCRIMINATION, AND RACISM 133 (John F. Dovidio & Samuel L. Gaertner eds., 1986) (stating that stereotyping leads to the evaluation of a person in accordance with cognitive structures that incorporate "the perceiver's knowledge, beliefs, and expectations about a human group") (emphasis omitted); Thomas E. Nelson et al., *Irrepressible Stereotypes*, 32 J. EXPERIMENTAL SOC. PSYCHOL. 13, 14 (1996) ("By *stereotypes* we mean beliefs or expectations about the qualities and characteristics of specific social groups.") (emphasis in original) (citations omitted).

16. *See, e.g.*, Daniel M. Wegner & John A. Bargh, *Control and Automaticity in Social Life*, *in* 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY 446, 451-52 (Daniel T. Gilbert et al. eds., 1998).

17. *See* Price Waterhouse v. Hopkins, 490 U.S. 228, 250 (1989).

discretionary judgments about ongoing workplace performance.[18] Yet such judgments are the stock-in-trade of employing and managing a labor force at all levels of job complexity, and appear essential to the proper functioning of the workplace.[19]

B. Unconscious Bias as Disparate Treatment

The form of conduct described above fits comfortably within the paradigm that lies at the heart of Title VII jurisprudence: disparate treatment. Central to the concept of disparate treatment in discrimination law is the existence of a causal link between a person's group identity or group-based trait and an actor's response to that person. The existence of that link, however, does not require that the actor be aware of the connection. Where a workplace supervisor's judgment is influenced by a worker's race or sex without the supervisor's knowledge or even against his wishes, trait-based disparate treatment can still result. Because the employer or his agent lacks awareness of this influence, the disparate treatment can be said to be inadvertent or unconscious.

The notion of unconscious disparate treatment might seem to fly in the face of the accepted requirement that actionable disparate treatment be "intentional."[20] But that term can carry at least two possible meanings. It can narrowly denote a form of scienter—an actor's conscious awareness of his reasons for acting. But it can also be used more broadly to refer to a causal link between a mental influence (e.g., the

---

18. It is possible to imagine ways in which employers can be said to engage in unconscious disparate treatment through the use of strictly objective criteria. For example, an employer's decision to adopt or retain a neutral requirement with disparate impact might be unconsciously motivated by the desire to screen out women or minorities. This story would appear to require positing an unconscious desire to hurt or exclude—that is, some kind of unconscious negative motive or animus—rather than, as discussed above, a covert mechanism of cognitive processing that interacts with information about race or sex to bring about an injurious effect. The question of whether the existence of unconscious animus has any psychological reality or empirical support is beyond the scope of this Article, and the special case of unconscious "animus-based" selection of objective, neutral criteria is not its central concern. For a discussion of this type of unconscious bias, see Brest, *supra* note 1, at 14-15; Oppenheimer, *supra* note 1, at 900-17; and Strauss, *supra* note 1, at 960.

19. Heavy reliance on objective screening devices is most common at the point of hiring, which generates only a small minority of employment discrimination claims. *See* John J. Donohue III & Peter Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983, 1015-19 (1991). Once employees are on the job, appraisals and decisions as to promotion, termination, etc., almost always make some use of subjective criteria. *See* discussion *infra* Part II.C.1.

20. An intent requirement is also recognized for employment claims brought under the Civil Rights Amendments and for violations of the constitutional requirement of equal protection. *See, e.g.*, General Bldg. Contractors Ass'n v. Pennsylvania, 458 U.S. 375, 390-91 (1982) (outlining an "intent requirement" for § 1981 claims); *see also* Daniel R. Ortiz, *The Myth of Intent in Equal Protection*, 41 STAN. L. REV. 1105, 1105 (1989). In contrast, there is no intent requirement for making out a disparate impact claim, and disparate impact and disparate treatment claims are often distinguished on the ground that the latter require a showing of "intent" whereas the former do not turn on any mental element. *See* 1 BARBARA LINDEMANN & PAUL GROSSMAN, EMPLOYMENT DISCRIMINATION LAW 81-114 (ABA Section of Labor and Employment Law ed., 3d ed. 1996).

perception of race or sex) and the outcome of a decision.[21] On this causal view of the intent requirement, disparate treatment—treatment that differs "because of" race or sex—can occur either "on purpose" or inadvertently. A decisionmaker can be aware of the causal connection, but he need not be. It then becomes possible to speak of "unconscious disparate treatment" of employees in the workplace, and to characterize such conduct as "intentional" in the causal sense of that term.[22]

C. How Important Is Unconscious Disparate Treatment?

How often do employees fall victim to unconscious disparate treatment in the workplace? In the current state of knowledge of human psychology, it is impossible to provide a definitive answer to this question. The suggestion that decisions regarding employees in the workplace might be inadvertently contaminated by category-based biases or stereotypes is not based on the observed operation of such biases in real-life workplace settings—nor could it be. Rather, it is based on the results of highly artificial experiments conducted by research psychologists under controlled laboratory conditions. Such experiments often present subjects with very limited information about the "target"—that is, the person or performance to be

---

21. A great deal of semantic and conceptual confusion surrounds the concept of discriminatory intent in commentary and law, and this confusion infects discussions of unconscious bias. The term "intent" sometimes carries the connotation of transparent awareness or comprehension of one's own motives. On this usage, possessing an "intent" to discriminate can be construed as discriminating "deliberately" or "on purpose," and discrimination that is inadvertent can be described as "unintentional." The terminological confusion is compounded by usage in the tort context, where "intentional" is generally taken to mean purposeful or deliberate rather than, more broadly, causally related to some state of mind. See *supra* note 20 and *infra* note 22 for discussions of "intentional" and "unintentional" torts and discrimination law.

22. Whether this is the sense actually adopted under existing employment discrimination laws such as Title VII is a matter of controversy. *See* discussion *infra* Part II.A on the scope of existing workplace discrimination statutes in doctrine and practice.

The principal proponent of the broader, "causal" account of the concept of intent is David Strauss, who argues that "intentional" discrimination under current law does, and properly should, include all forms of disparate treatment, or different treatment "because of" race, whether produced knowingly or unknowingly. Strauss, *supra* note 1, at 1015. As Strauss recognizes, the notion of "intentionality" is capacious enough to encompass decisions or actions undertaken without awareness on the part of the perpetrator of the role of a protected trait in the decisionmaking process—so long as race or sex, for example, does in fact make a difference. *Id.* at 956-57. Strauss proposes the so-called "reverse the groups" test to determine if a protected trait "makes a difference" or is the "but-for" cause of a harm. *Id.* at 956-59. A person discriminates with "intent" if, *ceteris paribus*, he would have acted differently in the absence of the protected trait. *See id.* at 958.

As noted, "intentional" discrimination in the broader sense stands in contrast to actions that simply create a disparate outcome or impact. The fact that the term "intent" has been made to do service for the distinction between the presence or absence of self-awareness of motives, as well as between disparate treatment and disparate impact, has produced some ambiguity in the literature addressing the scope of the antidiscrimination principle. Thus, a commentator who argues for doing away with an "intent" element for discrimination appears to be arguing that all discrimination law should be analyzed using the disparate impact paradigm. *See, e.g.*, McGinley, *supra* note 1, at 1463-73; *see also supra* note 6 (discussing disparate impact).

assessed. These experimental schemes necessarily depart significantly from the far richer and more complex situations encountered in the real world.[23] Moreover, laboratory-based experiments hold few useful "take-home" lessons for employers seeking to avoid discriminatory behaviors. Some studies provide evidence for the negative influence of race or sex on interpersonal assessments. Others elicit a positive bias from subjects under some circumstances, which can enhance the prospects of some members of traditionally disfavored groups; yet others demonstrate no significant influence at all.[24] The observed responses are highly

---

23. *See* Lee J. Jussim, & Jacquelynne Eccles, *Are Teacher Expectations Biased by Students' Gender, Social Class, or Ethnicity?*, *in* STEREOTYPE ACCURACY: TOWARD APPRECIATING GROUP DIFFERENCES 246 (Yueh-Ting Lee et al. eds., 1995) ("The overwhelming majority of social psychological research on stereotypes has been experimental laboratory studies."). Those studies "often use artificial or impoverished social stimuli." *Id.* at 247. Moreover, the experiments are based on encounters that are "usually with a stranger, for a period of an hour or less. And, of course, the laboratory studies primarily use college students as subjects." *Id.*; *see also id.* at 260 (remarking on the difficulty of using controlled experimental observations to draw conclusions about everyday behaviors in social settings); Clark R. McCauley, *Are Stereotypes Exaggerated? A Sampling of Racial, Gender, Academic, Occupational, and Political Stereotypes*, *in* STEREOTYPE ACCURACY: TOWARD APPRECIATING GROUP DIFFERENCES, *supra*, at 218 (same) [hereinafter McCauley, *Are Stereotypes Exaggerated?*]; Clark R. McCauley et al., *Stereotype Accuracy: Toward Appreciating Group Differences*, *in* STEREOTYPE ACCURACY: TOWARD APPRECIATING GROUP DIFFERENCES, *supra*, at 293 (same) [hereinafter McCauley et al., *Stereotype Accuracy*]; Martell, *supra* note 14, at 1939, 1954 ("As with any laboratory study, issues of generalizability inevitably arise."); Swim et al., *supra* note 14, at 423 (observing that virtually none of the multiple studies reviewed on the influence of sex-based biases on judgments and behavior "matched the complexity of real-life hiring or job evaluations decisions").

24. *See, e.g.*, Diane Kobrynowicz & Monica Biernat, *Considering Correctness, Contrast, and Categorization in Stereotyping Phenomena*, *in* 11 ADVANCES IN SOCIAL COGNITION, *supra* note 15, at 113 (describing work showing that stereotypes sometimes favor seemingly disfavored groups); R.D. Arvey & K.R. Murphy, *Performance Evaluation in Work Settings*, 49 ANN. REV. PSYCHOL. 141, 157 (1998) (noting that "there is no scientific support for the opinion that the format or specificity of appraisal systems [(i.e., whether subjective or more objective)] has a significant effect on race or gender bias in performance ratings"); Monica Biernat & Melvin Manis, *Shifting Standards and Stereotype-Based Judgments*, 66 J. PERSONALITY & SOC. PSYCHOL. 5 *passim* (1994); Monica Biernat et al., *Stereotypes and Standards of Judgment*, 60 J. PERSONALITY & SOC. PSYCHOL. 485 *passim* (1991); Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5 *passim* (1989); Kent D. Harber, *Feedback to Minorities: Evidence of a Positive Bias*, 74 J. PERSONALITY & SOC. PSYCHOL. 622 *passim* (1988); Lee J. Jussim, *Social Perception and Social Reality: A Reflection-Construction Model*, 98 PSYCHOL. REV. 54, 65 (1991) ("Whether individual targets are men and women, Blacks and Whites, old and young, or upper class and lower class, perceivers generally judge them far more on the basis of their observable relevant personal characteristics than on their membership in these social groups.") (citing numerous studies); Ziva Kunda & Paul Thagard, *Forming Impressions from Stereotypes, Traits, and Behaviors: A Parallel-Constraint-Satisfaction Theory*, 103 PSYCHOL. REV. 284, 302 (1996) (asserting that "the evidence that stereotypes dominate impressions is inconclusive at best, and that there are reasons to believe otherwise"); McConahay, *supra* note 14, *passim*; Nieva & Gutek, *supra* note 14, at 268-69 (noting that some studies show "evaluation bias favoring males," others that "women receive[] disproportionately favorable praise compared to men given similar performance," and yet others that there is "no difference in the evaluation of males and females"); Swim et al., *supra* note 14, *passim* (also finding mixed results in sex bias studies); Tosi & Einbender, *supra* note 14, at 721.

variable and context specific and often differ in unexpected and unexplained directions. The effects of protected attributes often depend on small variations in the selection of experimental subjects or targets, or seemingly arbitrary modifications in experimental presentation or design.[25] Without the ability precisely to match real-life to laboratory conditions, it becomes difficult to use experimental observations to predict with any certainty the degree to which cognitive biases or stereotyping will influence actual decisionmaking in a particular employment setting.

In sum, little of practical value is known about the real-life prevalence or influence of the mental categories that sometimes appear to explain some behaviors in the experimental setting.[26] At best, experimental social psychology suggests the possibility that hidden cognitive biases or stereotyping might on occasion influence subjective decisionmaking in the workplace without in any sense proving that such "mental contamination" is pervasive or that it is operating, or operating in one direction, in any particular case.

But even if sex or race-based bias does occasionally infect the decisionmaking process, it does not follow that the bias will make any concrete difference to the outcome of an actual workplace decision or that it will result in adverse treatment or denial of benefits in most cases or in any given case in which it plays a part. If a decision against an employee is influenced by race, it could be influenced to a variable degree, because race could distort the application of a neutral criterion only a little or a lot. If subtracting the influence of race would change the outcome of a decision in a particular case, then race can be considered a "determinative" or "but-for" cause of the outcome.[27] But if eliminating the influence of race would not

---

See generally 23 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY (Mark P. Zanna ed., 1990) (reporting mixed results on experimental studies of stereotyping); 7 THE PSYCHOLOGY OF PREJUDICE: THE ONTARIO SYMPOSIUM (Mark P. Zanna & James M. Olson eds., 1994) (same); COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR (David L. Hamilton ed., 1981) (same).

25. *See* sources cited *supra* notes 23-24; *see also* Krieger I, *supra* note 1, at 1313 (There is evidence to suggest that "we cannot say that intergroup bias will *always* cause discrimination to occur, nor can we predict exactly *when* discrimination will occur. . . . In short, we cannot expect systems in which cognitive or other situation-sensitive forms of bias are operating to present neat, consistent patterns.") (both emphases in original).

26. *See, e.g.*, Nelson et al., *supra* note 15, at 35 (concluding from a review of the literature that "social psychology is in a poor position to make any definitive assertions about the overall 'level' or 'degree' of stereotyping in everyday life").

27. The terminology is Stonefield's. Sam Stonefield, *Non-Determinative Discrimination, Mixed Motives, and the Inner Boundary of Discrimination Law*, 35 BUFF. L. REV. 85, 94 (1986). To illustrate the difference, suppose that employees were evaluated for promotion by a committee of three supervisors and rated on a numerical scale. An employee is required to obtain a score of 100—as a sum of the individual supervisors' scores—to win a promotion. Suppose that a black employee's race caused each of the three supervisors inadvertently to give the employee 10 fewer points than the employee would have received if he had been white. In the absence of the unconscious bias, the employee would have scored 80. With the bias, he scores 50. The employee did not miss out on the promotion "because of" race because he scored too low either way. For additional discussion of how the law should treat differences in treatment that yield no differences in job outcome, see Ernest F. Lidge III, *The Meaning of Discrimination: Why Courts Have Erred in Requiring Employment Discrimination Plaintiffs to Prove That the Employer's Action Was*

in fact make any difference to the assignment of tangible workplace benefits or rewards (even if it skews the process in some sense) then race could be considered "nondeterminative" of the outcome. Thus, even if most supervisors are often unconsciously biased against minority or female employees, it does not follow that those biases routinely result in disparate treatment in the assignment of tangible or measurable benefits, or even that they frequently do. The extent to which inadvertent category-based biases are generally determinative or nondeterminative of workplace outcomes cannot be known with any degree of certainty. But what is known about the nature of "mental contamination" suggests that such biases may often be partial or "subtle" in their effects. This suggests they may play a part in decisions only intermittently and rise to the level of determining concrete outcomes even less often.[28]

As the discussion indicates, the uncertainties surrounding the questions of whether and how often group-based cognitive biases come into play or produce tangible harms will prove important in determining whether imposing liability for this type of conduct is likely to optimize social welfare. As argued below, a system of individualized liability that seeks to identify hidden biases that may only infrequently determine outcomes will yield few effective precautions against those biases, will encourage inefficient overinvestments in precautionary measures, and will tend to generate large errors in the award of compensation.

D. Rational and Irrational Unconscious Bias

For the purpose of the ensuing analysis, it is helpful to draw a distinction between two general categories of unconsciously influenced actions in the workplace. The categories track a distinction made in discrimination scholarship between "rational" or "statistical" disparate treatment, and discrimination that is the manifestation of "irrational" prejudice.

If average productivity correlates with group membership, then the race or sex of an employee can be a form of useful, if quite imperfect, information about the current or projected job performance of that person. The information may be less useful than more specific and individualized information, but the latter may be difficult to obtain and interpret. If information is costly and stereotypical assumptions are roughly accurate for the group as a whole, the employer may be able to select a more productive workforce more cheaply by relying on race or sex-based generalizations.[29] But the use of such "statistically valid" information in

---

Materially Adverse or Ultimate, 47 U. KAN. L. REV. 333 (1999).

28. *See infra* Part II.C.1.

29. For a recent summary of the concept of statistical discrimination with citations to the economics literature, see Kenneth J. Arrow, *What Has Economics to Say About Racial Discrimination?*, 12 J. ECON. PERSP., Spring 1998, at 91, 96-97; *see also* David A. Strauss, *The Law and Economics of Racial Discrimination in Employment: The Case for Numerical Standards*, 79 GEO. L.J. 1619, 1622-23 (1991).

There is a growing body of work in cognitive psychology examining the uses of so-called "accurate" stereotyping—that is, cognitive categories or generalizations that reflect factually correct group differences. *See, e.g.*, STEREOTYPE ACCURACY: TOWARD APPRECIATING GROUP DIFFERENCES, *supra* note 23; Kobrynowicz & Biernat, *supra* note 24, at 119-123; Jussim, *supra*

workplace decisionmaking need not be conscious. Rather, it is possible for "rational" discrimination to take the form of unconscious stereotyping or inadvertent bias. An employer may take the most efficient course without meaning to do so.

On the other hand, trait-based biases that lead to the arbitrary discounting or skewed application of information pertinent to productivity may result in errors or mismatches in the selection and management of personnel—mismatches that could be avoided if bias did not operate. Then racial biases or sexual stereotypes might get in the way of choosing the "best" or most productive employees or of accurately evaluating employee performance based on the information otherwise at hand. In that case, a personnel process administered by unconsciously biased supervisors would do a worse job of assessing employees than one that used supervisors unaffected by those cognitive biases. Generally speaking, bias could be considered irrational if there exists any alternative process for evaluating employees that is both more neutral (in that the influence of unconscious trait-based biases would be diminished) as well as more cost effective (in that the balance of input and output costs in administering the alternative personnel system would be more favorable to the firm). The alternative processes might, but need not, make use of greater amounts, or different types, of information about employees.[30] In sum, bias is rational if the use of trait-based generalizations helps the employer make a better or more cost effective choice than could be made without relying on a stereotypical assumption.[31] In contrast, irrational bias involves the use of trait-based information to make a less desirable decision than could otherwise be made by eliminating the categorical bias.

---

note 24, at 69.

30. The adoption of a particular personnel system could improve the position of the firm overall without necessarily resulting in the retention or promotion of more productive employees. That is because some highly productive employees might be quite costly to identify and obtain. If slightly less productive employees could be identified and retained much more cheaply, the employer might come out ahead despite a decline in individual productivity. *See, e.g.*, MARK KELMAN & GILLIAN LESTER, JUMPING THE QUEUE: AN INQUIRY INTO THE LEGAL TREATMENT OF STUDENTS WITH LEARNING DISABILITIES 199-208 (1997) (distinguishing between concepts of gross productivity of employees—which does not take the costs of selecting, sorting and managing those employees into account—and net productivity, which does); Mark Kelman, *Concepts of Discrimination in "General Ability" Job Testing*, 104 HARV. L. REV. 1158, 1198 (1991). Kelman suggests that employers are concerned only with net productivity. *Id.* at 1198-1204.

31. A decisionmaking method that is best for the employer need not necessarily be best for society as a whole. It has been argued that "rational" or statistically valid discrimination, although generating positive benefits for the employer, produces socially undesirable "negative externalities" and should be suppressed. *See, e.g.*, Stewart Schwab, *Is Statistical Discrimination Efficient?*, 76 AM. ECON. REV. 228, 228 (1986); Cass R. Sunstein, *Three Civil Rights Fallacies*, 79 CAL. L. REV. 751, 758-59 (1991); Cass R. Sunstein, *Why Markets Don't Stop Discrimination*, 8 SOC. PHIL. & POL'Y, Spring 1991, at 22, 26-29 [hereinafter Sunstein, *Why Markets*]; *see also infra* note 194 (reviewing arguments that the goal of antidiscrimination law should be to eliminate both "irrational" as well as statistically valid forms of disparate treatment, because both reduce overall social welfare).

"Rational" unconscious bias may persist because it is efficient for the employer.[32] Although there is no definitive proof of the operation of "irrational" forms of unconscious bias in real-life profit-making enterprises, it is nevertheless possible to give some account of how such irrational biases might arise and persist. Based on experimental results, some cognitive psychologists have suggested that vulnerability to "mental contamination" from categorical assumptions and stereotypes is an intrinsic structural feature of human information processing and social judgment. But why might people retain these seemingly dysfunctional habits of thought? Theorists have posited the "cognitive miser" model to explain the development and habitual application of broad categorical assumptions to social judgments.[33] The mind has evolved methods that strike an overall compromise between the costs and benefits of attending to the full array of individualized information available in multiple social encounters. The repeated use of "quick and dirty" mental rules of thumb may function most efficiently in the aggregate, but may not represent the best method case by case. Although producing the best results in the range of social situations encountered overall, the rules may "reflect less than optimal information processing" in an individual interaction or for a specialized purpose.[34] Specifically, entrenched habits of thought might produce less than optimal management of a labor force. Then a system that indulged those habits could be considered irrational if a more neutral evaluation system—that is, one less influenced by category-based cognitive biases—would lower net costs to the firm.[35]

This analysis suggests that eliminating some types of unconscious group-based biases in the workplace would be a desirable result from the employer's as well as society's point of view. A rational firm would therefore choose to reduce or

---

32. *See* discussion *infra* text accompanying note 35.

33. *See* Philip E. Tetlock, *Accountability and Complexity of Thought*, 45 J. PERSONALITY & SOC. PSYCHOL. 74, 74 (1983) ("A principle of least effort seems to guide much human judgment and decision making."); *see also* Brown et al., *supra* note 1, at 1495 (discussing the cognitive miser model of cognition); Neil Macrae et al., *Stereotypes as Energy-Saving Devices: A Peek Inside the Cognitive Toolbox*, 66 J. PERSONALITY & SOC. PSYCHOL. 37, 37 (1994) (same); Bargh, *supra* note 11, at 362 (discussing the cognitive miser concept and explaining that "[t]he reliance on simple decision rules and on pigeonholing of individuals into stock characters or categories was viewed mainly as a matter of strategic necessity, or even as an adaptive way of dealing with our mental shortcomings as human beings").

34. *See* Hamilton & Sherman, *supra* note 15, at 55-56.

35. It could be argued that unconsciously biased decisionmaking cannot be regarded as "irrational" so long as eliminating the distortions in decisionmaking is extremely expensive or impossible in practice. Since costs are costs, it should not matter whether the costs of eliminating the use of group identity in the employment calculus would take the form, as with statistical discrimination, of depriving the decisionmaker of valuable information or, as in the "irrational bias" case, of purging the evaluator's mind of dysfunctional but entrenched cognitive habits. But regardless of whether the practical inability to accomplish the latter result is properly regarded as a form of "cost" to the firm, there remains an important and useful—if theoretical—distinction between statistical discrimination and irrational unconscious bias: the first uses race-based information to help make a better choice than could be made without it, while the latter uses race-based information to make a worse choice than could otherwise be made without it regardless of whether such reduced reliance is psychologically attainable in fact. Put another way, if the use of group-based information could feasibly be eliminated *without cost*, the employer in the latter case would choose to do so, but in the former case would not.

eliminate the influence of group-based biases on their personnel process if such a result could be effectively achieved. In effect, the firm might choose to "reprogram" its supervisors' brains to eliminate some stereotyped information processing or otherwise to restructure its evaluation system to minimize the influence of bias if such changes were feasible and not so costly as to wipe out the gains from the resulting improvements. These observations raise the question of whether, and how, such group-based biases, if in fact present, could be eliminated. Subsequent sections of this Article will take up this question.

E. Discrimination as Accident

The preceding Section describes how mental processes that psychologists have theorized are part of normal cognitive functioning could possibly give rise to unconscious disparate treatment in the workplace and could cause "irrational" unconscious bias to persist. This Section considers how unconscious disparate treatment can be considered a kind of workplace accident. An accident has been defined as a "harmful outcome[] that neither injurers nor victims wish[] to occur."[36] Perhaps the most important way in which unconscious discrimination is like an accident is that it creates an "unintended" risk of a particular type of costly harm as a byproduct of an economically useful activity—that is, the employment of workers in the service of some productive enterprise.[37]

As explained more fully below, race or sex discrimination in the workplace can be regarded as a type of tortious accident. But discrimination is a peculiar type of tort. Any adverse action taken by employer against employee, regardless of the employers' reasons, can be considered injurious to the employees' interests as such. But under the framework created by existing laws against workplace discrimination, not all such injuries are actionable. Laws such as Title VII create no generalized duty to treat employees favorably, but only to refrain from treating employees differently "because of" race or sex.[38] As already suggested, that duty can arguably be breached if the difference in treatment is consciously (that is, intentionally) or unconsciously motivated by the forbidden characteristic. The "tort" of unconscious discrimination is not an "intentional" one—it is not committed for the purpose of bringing about the harm—but rather is an accidental tort that is the unintended byproduct of an otherwise useful activity.

---

36. STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW 1 (1987). Although the term "accident" is ubiquitous in liability scholarship, commentators rarely define the term. *See, e.g.*, GUIDO CALABRESI, THE COSTS OF ACCIDENTS (1975) (providing no definition).

37. Here the term "intent" is employed more narrowly than in discrimination law. The category of "intentional tort" is reserved for injuries that are inflicted deliberately and knowingly, in that a person takes action with the desire or wish to inflict harm. *See* discussion *supra* text accompanying notes 21-23 and *infra* text accompanying notes 42, 54.

38. *See infra* text accompanying notes 59, 62, 64 (discussing breach of duty, compensable harm, etc. in a discrimination context).

II. LEGAL RESPONSES TO UNCONSCIOUS DISPARATE TREATMENT

Before proceeding to the analysis of whether liability for unconscious or "accidental" disparate treatment would produce a socially optimal result, it is important to consider the preliminary issue of whether, and how, existing laws governing workplace discrimination currently address the phenomenon of unconscious disparate treatment.

A. Does Current Law Cover Unconscious Disparate Treatment?

Does Title VII already provide a cause of action for the type of unconscious disparate treatment described above? Some scholars have suggested that there is nothing in the letter of Title VII that rules out imposing liability for this type of conduct.[39] As already noted, the statutory language forbidding discrimination "because of" protected traits is arguably ambiguous enough to encompass both deliberate and inadvertent forms of causation,[40] and the doctrinal requirement that the employer harbor "intent" to discriminate could be construed as capacious enough to cover actions inadvertently triggered by protected traits.[41] Nonetheless, there is considerable ambiguity in the case law surrounding the scope of the "intent" requirement for claims alleging disparate treatment. There thus remains some doctrinal uncertainty as to whether "intentional" discrimination encompasses unconscious as well as conscious "motives" for action.[42]

The most suggestive evidence that Title VII doctrine has evolved towards formally excluding recovery for inadvertent disparate treatment comes from the operation of the *McDonnell Douglas* evidentiary formula in individual disparate treatment claims.[43] Under *McDonnell Douglas*, a plaintiff can create a presumption of

---

39. *See, e.g.*, Brodin, *supra* note 9, at 987-97; Deborah C. Malamud, *The Last Minuet: Disparate Treatment After* Hicks, 93 MICH. L. REV. 2229, 2237 (1995); McGinley, *supra* note 1, at 1463-73; Strauss, *supra* note 1, at 937-39; D. Don Welch, *Removing Discriminatory Barriers: Basing Disparate Treatment Analysis on Motive Rather Than Intent*, 60 S. CAL. L. REV. 733, 734-36 (1987).

40. The statute itself also provides remedies for "intent" to discriminate, *see* Civil Rights Act of 1964 § 706(g), 42 U.S.C. § 2000e-5(g) (1994), although it does not define that term.

41. See *supra* text accompanying note 22 for a discussion of David Strauss's work.

42. Linda Krieger argues that courts have effectively narrowed the scope of the "intent" requirement under Title VII, both in doctrine and practice, to cover only deliberate acts of discrimination. Krieger II, *supra* note 1, at 1164 (arguing that the bulk of unconscious disparate treatment in the workplace goes unremedied because courts have routinely "constructed" and applied antidiscrimination statutes in a manner that, although "sufficient to address the deliberate discrimination prevalent in an earlier age, is inadequate to address the subtle, often unconscious forms of bias that Title VII was also intended to remedy") (footnote omitted); *see also* Chamallas, *supra* note 1, at 467 (implying that existing law encompasses only deliberate and not unconscious disparate treatment).

43. McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).

discriminatory motive by establishing certain facts.[44] The employer is then required to come forward with a nondiscriminatory reason for the action taken.[45] The trier of fact must then decide whether the reason given was the "true" reason for the employer's decision, or whether that reason is "pretext"—that is, not the "real" reason for the decision.[46] Alternatively, in so-called "mixed motive" cases, the trier must decide whether, even if there is evidence that the employer relied on both trait-based and non-trait-based reasons for the decision, the outcome would have been the same in the absence of discriminatory intent. If the same decision would have been forthcoming, the plaintiff is not entitled to compensatory or equitable relief.[47]

The *McDonnell Douglas* formulation is clearly geared to a narrow view of discriminatory intent: its operation depends on a defendant's possessing a conscious or deliberate state of mind. *McDonnell Douglas* demands that the defendant supply reasons or "motives" for a decision or action taken in the employment setting. The requirement that the actor actually articulate his reasons for action assumes that the actor is fully aware of why he acted. The idea of pretext—in which an actor tries to cover up his "true" reasons by supplying false ones—presupposes that the actor's reasons are transparent to him: he can and does in fact know through introspection what his reasons *really* are. If the plaintiff is charged with showing that the explanation the defendant provides is "false," this suggests that the defendant did not have the stated reason(s) in mind when he acted, but rather some "invidious" reason (e.g., race or sex). In effect, the final step in *McDonnell Douglas* turns on questions of sincerity and credibility: Is the defendant lying about his own motives for action, or is he not?[48]

---

44. These are: that the employee was qualified for a position or benefit; that he was denied it; and that the position remained open or was filled by a person from another group. *See id.* at 802.

45. *See id.*

46. *See id.* at 804. Once the finding of pretext is made, the trier of fact is permitted—but after the Supreme Court's decision in *St. Mary's Honor Center v. Hicks*, 590 U.S. 502 (1993), is not required—to infer from the evidence that the employer acted with discriminatory intent. *See* Malamud, *supra* note 39, at 2234 (discussing the *Hicks* opinion).

47. *See, e.g.*, Price Waterhouse v. Green, 490 U.S. 228, 258 (1998). Prior to the Civil Rights Act of 1991, Title VII was ambiguous on the placement of burdens of production and proof for the elements of a dual or mixed motive case. The courts had construed the statute to require the defendant to bear the burden of showing that discriminatory factors made no difference to any employment-related decision, and such a showing negated liability. *See id.* at 248. The statute was amended in 1991 expressly to permit a finding of liability (and possibly an award of attorney's fees) if the plaintiff shows that discrimination was a "motivating factor" in the decision. *See* Civil Rights Act of 1964 § 703(m), 42 U.S.C. § 2000e-2(m) (1994), *amended by* Civil Rights Act of 1991, Pub. L. No. 102-166, § 107(a). No remedy is forthcoming, however, if the employer proves that the "same action" would have been taken even absent the discriminatory motive. *See* Civil Rights Act of 1964 § 706(g), 42 U.S.C. § 2000e-5(g) (1994), *amended by* Civil Rights Act of 1991, Pub. L. No. 102-166, § 107(b).

48. The fact-finder (which can be judge or jury under current law) is not required to find that there was discrimination even if it determines that the proffered reason was pretext, since the defendant may have had other nondiscriminatory reasons for an action. *See St. Mary's Honor Ctr.*, 509 U.S. at 509-11. But the point is that *McDonnell Douglas* assumes that the question of discrimination turns only on reasons or motives of which the actor is aware, and on none of which he is unaware.

The *McDonnell Douglas* framework seems all wrong for the task of getting at hidden cognitive processes. The requirement that defendants give reasons and the very idea of pretext itself are predicated on the assumption that persons are fully cognizant of the motives for their actions.[49] Moreover, the idea of a "pretext" is cartoonish: the decisionmaker has a single motive, which is either suspect or not. He gives a reason, and he is either lying or not. The introduction of the possibility of unconscious motivation shows the inadequacy of this framework. To the extent that the factors that influence human decisionmaking can be identified as "reasons" or "motives,"[50] the reality of human decisionmaking is that multiple "reasons" underlie every decision in which unconscious processes are at work, and those "reasons" can take forms that are quite obscure to the actor. The crude polarity between the conscious and unconscious fails to capture the complex gradation of individuals' awareness of the antecedents of their decisions, which can represent a tangle of factors that are half-conscious and partly understood.[51] Supervisors may forget, or be oblivious to, many of the reasons they acted in the first place. Or they may simply be influenced by cognitive mechanisms of which they are unaware.

Because information about race can operate to distort the application of otherwise "neutral" criteria, the concept of "pretext"—which goes to what the employer *thinks* he is doing—is conceptually irrelevant to whether unconscious bias is at work.[52] The supervisor may have a perfectly sincere and valid reason in mind for a decision, and yet still be acting "because of" race. In the typical case in which unconscious bias infects the process, race is rarely a "sole cause," but rather operates to distort or "skew" the application of other legitimate factors (or neutral "reasons") that go into workplace evaluations. For example, the race of the employee may cause the employer to place more weight on one otherwise legitimate aspect of the employee's performance, and less weight on another aspect, than if the employee were of a

---

49. See, for example, Krieger II, *supra* note 1, at 1164-66, for a critique of the assumptions of "transparency" and rationality of motives that underlie the pretext and mixed motive models in antidiscrimination law; David N. Rosen & Jonathan M. Freiman, *Remodeling* McDonnell Douglas*:* Fisher v. Vassar College *and the Structure of Employment Discrimination Law*, 17 QUINNIPIAC L. REV. 725, 761-63 (1998) (complaining that the *McDonnell Douglas* formulation ignores the fact that "in many cases it is neither useful nor possible to distinguish among employers who rely on prohibited stereotypes based on the level of their conscious awareness of having done so").

50. For discussions of the use of the concept of "motive" in antidiscrimination law and jurisprudence, see, for example, Susan Bisom-Rapp, *Of Motives and Maleness: A Critical View of Mixed Motive Doctrine in Title VII Sex Discrimination Cases*, 1995 UTAH L. REV. 1029, 1030-34 (exploring how the mixed motive doctrine fails to capture the social reality of working women); Brodin, *supra* note 9, at 987-97 (arguing that a restrictive definition of intent is not consistent with the legal system's approach and does not facilitate Title VII policy); Paul J. Gudel, *Beyond Causation: The Interpretation of Action and the Mixed Motives Problem in Employment Discrimination Law*, 70 TEX. L. REV. 17, 71-82 (1991) (analyzing the mixed motive concept in law and arguing that the *Price Waterhouse* approach is fundamentally flawed); Welch, *supra* note 39, at 736-40 (arguing that motive rather than intent should be controlling in Title VII cases).

51. *See generally* Wegner & Bargh, *supra* note 16.

52. It may not be irrelevant as a matter of evidence, however. *See infra* text accompanying note 225.

different race. The exaggerations or deviations can only be measured against the decisionmaker's responses under hypothetical "baseline" or counterfactual conditions—the conditions that would obtain if the employee were of a different race or if the cognitive biases were absent.

The description of unconscious bias as "skewing" the application of a valid set of decisionmaking criteria suggests that most decisionmaking in the employment setting arguably fits better into the "mixed motive" than the "pretext" category of discrimination claims. When unconscious biases are at work, it can be said both that the decisionmaker possesses a "valid" reason for action and also that the protected trait was a "motivating factor" in the decision. But the notion of mixed motives, as it has been articulated in cases and commentary, is not a very good fit either: people engaged in evaluating others in the workplace do not generally possess distinct reasons or motives that run along parallel lines and simultaneously "cause" a decision or contribute to the cause. The schematic "two-track" image created by mixed motive analysis fails to capture the operation of unconscious bias in social judgment. It is closer to psychological reality to say that employers or their agents apply a set of neutral and often reasonable criteria, but apply them imperfectly, erratically, inconsistently, or, at times, just plain differently to employees from protected categories.[53]

The foregoing discussion shows that current doctrine is formally at odds with liability for unconscious forms of disparate treatment. But what about actual cases? Does *McDonnell Douglas* or the current treatment of mixed motive cases stand as an important practical obstacle to the bringing and winning of claims arising from unconscious disparate treatment? Are there other important impediments to the prosecution of unconscious disparate treatment claims under current law?[54]

---

53. Indeed, this redescription upsets a number of core assumptions about the psychology of discrimination that underlie current doctrine in general, and the *McDonnell Douglas* paradigm in particular. As Linda Krieger points out, not only are motives assumed to be transparent, but decisionmakers are assumed to be consistent in their biased behavior, because discrimination is believed to spring from stable tastes or *ex ante* preferences, rather than from mental schemas with complex, variable, and context-dependent applications. Krieger I, *supra* note 1, at 1310 (criticizing as fundamentally at odds with the psychology of unconscious bias the "same actor" doctrine, which holds that "if the same person who hired an employee makes the decision to fire him, a strong inference or presumption of non-discrimination arises").

In a similar vein, it is important to recognize that the "irrational" cognitive generalizations that represent a response to protected traits such as race and sex are not sharply distinct from other mental schemas, but lie on a continuum with other decisionmaking processes. Generalizations based on employee characteristics other than those specifically covered by antidiscrimination laws may trigger deviations from rationality or consistency in some circumstances. Thus, the fact that a supervisor's decisionmaking practices do not perfectly correspond to the "reasons" given for action, or are not otherwise explicable on the basis of some employment-related goal, does not necessarily mean that the supervisor is acting on the basis of a protected trait. *See infra* text accompanying notes 136-40 (discussing the possibility that employment decisions often deviate from rationality in ways unrelated to protected traits).

54. Krieger also claims that, apart from the effects of the application of the *McDonnell Douglas* framework, there are other important impediments to the prosecution of unconscious discrimination claims. She argues that courts are hostile generally to unconscious discrimination claims because they assume that the "intent" requirement entails the need for deliberate or conscious animus. *See* Krieger II, *supra* note 1, at 1161-73. According to Krieger, "the entire

Although *McDonnell Douglas* may make it somewhat harder for victims of unconscious bias to win their cases, there is no reason to think that *McDonnell Douglas* serves to shut out all claims that are grounded in subliminal bias. That is because, as a practical matter, there inevitably will be some degree of overlap in the evidence that tends to prove "pretext" and the evidence that tends to prove unconscious bias. That evidence will often take the form of inconsistencies or disparities in treatment of similarly situated members of different groups. When those disparities are unexplained or poorly explained, some fact-finders will infer "deliberate" discrimination. The overlap in proof may cause some fact-finders to find pretext (erroneously, to be sure) when, in fact, the defendant's conduct was driven by inadvertency. It may be, however, that the types of stark disparities that deliberate animus can generate will prove quite rare when motives are unconscious.[55] The quality and quantity of comparative evidence for "pretext" that is generally available when unconscious bias operates may make for weaker cases and fewer holdings in favor of plaintiffs.

This account assumes that the fact-finder will take the requirement that it find "pretext" seriously. If it does not—if the fact-finder is impressed by unexplained disparities in treatment with little attention to the form of the motive—then the result will be that liability will sometimes attach for unconscious disparate treatment. On the other hand, if pretext is taken seriously, *McDonnell Douglas* might impede

---

normative structure of Title VII's injunction 'not to discriminate,' rests on the assumption that decisionmakers possess 'transparency of mind'—that they are aware of the reasons why they are about to make, or have made, a particular employment decision." *Id.* at 1185. For Krieger, the most telling evidence for the general acceptance of the "assumption of decisionmaker self-awareness" is the "moment of decision" language from Justice Brennan's *Price Waterhouse* plurality opinion. *Id.* (citing Price Waterhouse v. Green, 490 U.S. 228, 250 (1988)). Justice Brennan states:

> In saying that gender played a motivating part in an employment decision, we mean that, if we asked the employer at the moment of the decision what its reasons were and if we received a truthful response, one of those reasons would be that the applicant or employee was a woman.

Price Waterhouse, 490 U.S. at 250. As Krieger points out, this language only makes sense on the assumption that "employment decisionmakers have ready access to the workings of their own inferential process." Krieger II, *supra* note 1, at 1185. *But see* Rosen & Freiman, *supra* note 49, at 765-67 (suggesting an alternative construction of Brennan's remarks that is more consistent with coverage of unconscious stereotyping).

Apart from her reliance on Justice Brennan's remarks in *Price Waterhouse*, however, Krieger's evidence for a judicially imposed "conscious intent" requirement is remarkably thin. Discussions in cases and commentary on the requirement of showing "intent" appear to have in mind the distinction between disparate treatment claims (which turn on state of mind) and claims of disparate impact (which do not). *See* discussion *supra* note 6 (on disparate impact); *see also* General Bldg. Contractors Ass'n v. Pennsylvania, 458 U.S. 375, 388-91 (1981) (holding that § 1981, in requiring proof of "intent," does not contemplate disparate impact claims). Thus, the argument against an "intent" requirement is often an argument in favor of expanding on liability for practices with a disparate impact. *See, e.g.*, McGinley, *supra* note 1, at 1463-73 (stating that the "intent" requirement is an impossible burden for plaintiffs to meet and advocating the adoption of a negligence standard); Strauss, *supra* note 29, at 1644-46, 1654-56 (criticizing the intent standard and suggesting that it be replaced with a system where employers are fined if they do not hire minority workers in proportion to their representation in the population as a whole).

55. *See infra* text accompanying note 135.

plaintiffs' recovery in some cases in which unconscious bias is operating by shifting the focus away from objective evidence towards an evaluation of the sincerity or credibility of the defendant. If fact-finders take the concept of pretext seriously, they may be willing to overlook significant disparities in treatment or erratic employer behavior if convinced of a supervisor's sincerity in denying that "race was the reason."[56] How much of an obstacle *McDonnell Douglas* poses in individual cases is thus an empirical question that depends on how the *McDonnell Douglas* framework actually plays out on the facts of each case.[57]

The doctrines and practices that have evolved in cases said to involve "mixed motives" are, if anything, even less hospitable to establishing liability based on unconscious forms of bias than the conventions surrounding pretext cases. The "mixed motive" paradigm usually comes into play when there is some evidence supporting a permissible justification for the action taken against an employee. Courts then permit the case to go forward only if the plaintiff can supply some form of "direct" or "anecdotal" evidence that the employer may have acted for discriminatory reasons as well as legitimate ones. This demand would often rule out a mixed motive analysis for cases stemming from unconscious bias, because the kind of evidence that is generally available when motives are unconscious—bare, unexplained disparities in group treatment—would not ordinarily satisfy the courts' threshold standard. The types of anecdotal or direct evidence that courts customarily demand in these circumstances would be available almost exclusively where the discrimination is self-conscious or deliberate.[58]

---

56. Linda Krieger provides a neat illustration of how *McDonnell Douglas* might make it harder to prove unconscious discrimination in the face of evidence of disparate outcomes by making a case turn on the credibility of the defendant's account of why he acted. An ethnic minority worker is disciplined and fired. He sues for discrimination. Krieger constructs the following hypothetical sequence: The supervisor denies discrimination and points to the employee's specific transgressions. Evidence is introduced concerning the treatment of other workers in the plant, and there is "a subtle, yet discernible pattern of differential treatment emerging from the time records and personnel files obtained in discovery." Kreiger II, *supra* note 1, at 1162-63. Krieger contends that, despite these demonstrable overall disparities in treatment, the defendant prevails because the trier of fact is unconvinced that the differential treatment was consciously imposed because of the worker's ethnicity. Despite the plaintiff's efforts to prove otherwise, the fact-finder simply refuses to believe that "the plant manager was a racist and a liar." *Id.* at 1163. Since pretext has not been proved, the fact-finder finds no discrimination. *See id.*

57. Moreover, *McDonnell Douglas* is only an important factor in individual disparate treatment claims. Group claims—whether class actions or pattern and practice claims—do not stand or fall on "pretext," and *McDonnell Douglas* plays little part in the order of proof of these cases. *See* 1 LINDEMANN & GROSSMAN, *supra* note 20, at 44-47; Maurice R. Munroe, *The EEOC: Pattern and Practice Imperfect*, 13 YALE L. & POL'Y REV. 219, 247 (1995). In the end, group cases often turn on the cogency of the statistical proof offered by the parties, and the plausibility of the alternative theories for rationalizing numerical patterns. Because bias that operates unconsciously can sometimes—if not always—generate statistical evidence similar to that available for discrimination that is practiced "on purpose," it might be possible for some victims of unconscious discrimination to prevail in pattern or practice claims alleging disparate treatment. *See infra* text accompanying note 125 (discussing proof).

58. The circuits are split on the issue of what evidentiary standard should apply to mixed or dual motive cases. Some circuits have adopted a strict "direct evidence" requirement. This is generally construed to require direct proof, without inference, that the agent who made the decision

In sum, although there are no absolute statutory impediments to plaintiffs' recovery for unconscious forms of bias under current law, current doctrine and practice are stacked against recovery in many cases. This suggests that, as things now stand, employees victimized by unconscious forms of bias will only occasionally succeed under Title VII. This suggests as well that addressing unconscious disparate treatment more comprehensively might well require a significant extension or overhaul of the current legal regime. Specifically, either Title VII would have to be revised expressly to recognize unconscious disparate treatment claims, or doctrine and evidentiary practice would have to be reformed to permit the more effective prosecution of such claims.

B. Liability for Unconscious Disparate Treatment: Strict Liability or Negligence?

This Article has introduced the idea that the effects of unconscious bias can be viewed as a form of workplace accident. The proposal under consideration could be viewed as one to reform the existing liability regime for employment discrimination to more consistently assign the costs of these accidents to the employer. This proposal can be systematically assessed by looking to the three principal goals of a liability system for accidental harms: deterrence, compensation, and insurance. As a first step in investigating whether the proposed reform would further these objectives, this Section will discuss some preliminary considerations relevant to the optimal design of a liability rule to deal with unconscious bias—specifically the choice between strict liability and negligence regimes.

In the discrimination context, the elements going to liability—duty, breach of duty, and causation—are established by statute. Under the federal antidiscrimination statute, Title VII, for example, the employer owes a duty to the employee to refrain from taking any adverse action "because of" race or some other protected characteristic. The duty as well as the causal relationship are defined as a function of the actor's reasons or motives. The employer is permitted to "injure" or "harm"[59]

---

harbored discriminatory animus. *See, e.g.*, EEOC v. Wiltel, Inc., 81 F.3d 1508, 1515 (10th Cir. 1996); Langley v. Jackson State Univ., 14 F.3d 1070, 1075 (5th Cir. 1994); Jackson v. Harvard Univ., 900 F.2d 464, 467 (1st Cir. 1990). Several other circuits allow the use of "circumstantial evidence," but still require that the evidence tend to prove discriminatory animus on the part of the person involved in the contested employment decision. *See, e.g.*, EEOC v. Pape Lift, Inc., 115 F.3d 676, 684 (9th Cir. 1997); Fields v. New York Office of Mental Retardation & Dev. Disabilities, 115 F.3d 116, 122 (2d Cir. 1997); Hook v. Ernst & Young, 28 F.3d 366, 373-74 (3d Cir. 1994). Several circuits have stated outright that "statistical evidence by nature does not merit a mixed motive charge." Fuller v. Phipps, 67 F.3d 1137, 1143 (4th Cir. 1995) (citing Ostrowski v. Atlantic Mut. Ins. Co., 968 F.2d 171, 182 (2d Cir. 1992)); Griffiths v. CIGNA Corp., 988 F.2d 457, 470 (3d Cir. 1993) (quoting *Ostrowski*, 968 F.2d at 182); *see also Ostrowski*, 968 F.2d at 182 ("[P]urely statistical evidence would not warrant such a charge . . . ."). *See generally* 1 LINDEMANN & GROSSMAN, *supra* note 20, at 40-44.

59. *See supra* Part I.E (discussing adverse decisions in the workplace as generically "harmful"). An unfavorable action against an employee operates as a setback to the employee's interests, decreasing his "total utility" and making him worse off. However, not all "injurious" adverse actions inflicted by employers against employees are tortious or "actionable" harms under the employment discrimination laws. In this respect, discrimination differs somewhat in form from

the employee by taking adverse action against him (for example, by failing to hire, promote, or retain him, or by inflicting detrimental treatment of any kind) for other reasons or no reason at all, but not for reasons linked to protected characteristics.[60] Acting "because of" those factors gives rise to a breach of duty, and liability. The statute also prescribes remedies for the breach of the duty not to discriminate. Under Title VII, individual plaintiffs are entitled to back pay, equitable reinstatement, and in some cases front pay, compensation for emotional harms, and punitive damages. Group of claimants, or the EEOC in pattern and practice cases, can sometimes obtain more complex equitable remedies.[61]

Existing antidiscrimination laws effectively erect a strict liability regime: employers or their agents are liable for detrimental actions triggered by improper motives regardless of whether they exercise "due care."[62] Retaining the Title VII paradigm for unconscious disparate treatment amounts to imposing a type of strict liability for adverse actions against employees due to unconscious bias. Such a rule would take no account of the relative costs (to the enterprise) or benefits (to the enterprise and victim) of eliminating unconscious race-based bias in decisionmaking. Rather, the rule would be directed at fully charging the costs of the harms of discriminatory practices to the enterprise by forcing it to bear all costs of discrimination regardless of whether eliminating the harm is cost effective or whether the harms could in practice be eliminated.

The preceding discussion suggests that the first question to ask is whether strict liability is superior to negligence for unconscious disparate treatment claims.[63] If a negligence standard were adopted, courts would be faced with three tasks: fixing the standard of care, deciding whether the defendant has met that standard, and determining causation—that is, determining whether the adverse action against the

---

other types of workplace harms for which employers may be strictly liable, in that the employer's causal responsibility for the injury, without more, establishes liability. *See infra* text accompanying notes 67-70 (discussing the practical implications of this difference).

60. This statement is an oversimplification because common law and state law "just cause" rules create other exceptions. *See, e.g.*, J. Hoult Verkerke, *An Empirical Perspective on Indefinite Term Employment Contracts: Resolving the Just Cause Debate*, 1995 WIS. L. REV. 837. It also overstates the purity of strict liability in practice, since many strict liability regimes, such as products liability, effectively smuggle in elements of due care. *See infra* note 75.

61. *See* Civil Rights Act of 1964 § 706(g), 42 U.S.C. § 2000e-5(g) (1994) (remedies); 42 U.S.C. § 1981a(b) (1994) (punitive damages); *id.* § 1981a(b)(3) (1994) (compensation for emotional distress and mental anguish).

62. For a discussion of negligence and strict liability, see WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 85-122 (1987); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 163-80 (4th ed. 1992); SHAVELL, *supra* note 36, at 73-85. A rule of strict vicarious or enterprise liability applies to cases of ordinary disparate treatment. A different and more complex rule applies to sexual harassment claims, which are outside the scope of this Article. For a discussion of enterprise liability in the unconscious discrimination context, see *infra* Part II.C.4.

63. The concept of "negligent discrimination" is not unknown in legal scholarship, although the analyses so far have not made use of concepts of accidents law. *See, e.g.*, Brest, *supra* note 1; Oppenheimer, *supra* note 1; Strauss, *supra* note 1; Allen, *supra* note 1.

employee was taken "because of" a protected characteristic.[64] One important advantage of the strict liability rule is that it dispenses with the first two of these.[65] Since the first two inquiries are likely to be quite complex, cumbersome, and error-prone, the reduction from three to one will almost certainly result in a reduction in litigation and process costs and perhaps an increase in accuracy as well. One equally intensive and vexing factual inquiry remains, however: the mental element of causation. Did the employer base his decision on an impermissible factor—such as race or sex—or a permissible one?

The question of cause is harder to sort out in discrimination cases than in many other cases of liability arising from workplace harms. The important distinction here is between cases in which liability is effectively coextensive with causation by the agent (e.g., the employer), and cases in which liability requires establishing causation by the employer of a certain type. In the first type of case, establishing "internal" causation (i.e., causation internal to the workplace) establishes liability. Those are the cases in which strict liability almost always creates a huge advantage in simplicity and ease of adjudication. In the second type of case, it is not enough to establish that the employer was the agent of the injury, because the employer can bring about the same adverse result either innocently or tortiously; rather, it is necessary to distinguish between different types of internal causation. The advantage of strict liability is dissipated by the need to engage in the additional causal inquiry to distinguish between actionable internal causation and other kinds. For discrimination, this exercise is especially difficult because the distinction is grounded in a mental element. The fact-finder must determine whether the adverse action was taken "because of" a protected trait or not. That determination is unusually difficult in the case of unconscious disparate treatment.

As suggested, in most cases in which strict liability turns on the clean distinction between internal and external causation, it is often quite easy to show that the employer is responsible for the injury. There is no serious contention respecting the critical element of causation when a worker's arm is cut off by industrial machinery.[66] Establishing that the event happened on the job is enough to ground

---

64. This discussion neglects another element that is arguably relevant to deciding whether unconscious discrimination should be subject to a negligence rule: whether the harm inflicted by unconscious bias in the workplace is "reasonably foreseeable." *See, e.g.*, Stephen R. Perry, *Libertarianism, Entitlement, and Responsibility*, 26 PHIL. & PUB. AFF. 351, 356-57 (1997) (discussing the importance of the concept of foreseeability in assigning responsibility within liability schemes). As already suggested, and as discussed *infra* Part.II.C.2, unconscious disparate treatment is so elusive and difficult to demonstrate in the "real world" (as opposed to in a highly controlled laboratory setting) that there remains a serious question whether there is *any* unconscious disparate treatment in the workplace, let alone how much. It is thus unclear whether unconscious discrimination can be said to be "reasonably foreseeable" in any accepted sense of that term.

65. *See* POSNER, *supra* note 62, at 175-76. This is an oversimplification, because actual strict liability regimes often limit the scope of recovery to cases in which the injury was caused by a "defective" product—a limitation that effectively incorporates a categorical standard of care. But that limitation is not a necessary feature of any strict liability rule. *See supra* note 60.

66. Thus, most workplace risks that would expose employers to liability through workmen's compensation programs, or otherwise, are easy to monitor because there is, for all practical purposes, absolute liability for those risks: If the accident happened *in* the workplace—for example,

liability. But sometimes the external/internal causation distinction is quite difficult to make: the toxic torts cases are an example.[67] In such cases, liability depends on whether some toxic agent or condition to which the worker was exposed on the job, as opposed to some influence *outside* the workplace, is causally responsible for the injury. Making the distinction between external and internal cause is difficult in many toxic torts cases because factoring out the influence of external or background risk for many diseases, such as cancer, is not at all straightforward.[68]

The inquiry for workplace discrimination is more complicated still. As with toxic torts, there is always a serious possibility that an alternative, and nonactionable, set of influences is responsible for the adverse event. But for discrimination, all alternative causes, whether innocent or not, are *internal* to the workplace. There is never any question that the employer caused the "injury" in the sense of being responsible for the adverse event because only the employer can fire or discipline an employee. But because of the way in which the employer's duty is defined under the antidiscrimination laws, liability must implicate yet another "layer" of causation: the plaintiff must show that the employer's action is causally linked to the employee's race or sex. Cases alleging unconscious discrimination in the workplace thus bear an important resemblance to toxic torts claims in that causation is always a central issue. There is always the possibility that an alternative and nonactionable set of influences is responsible for the adverse event. But, unlike with toxic torts, the causal question is not whether the employer caused the adverse event, as opposed to some outside influence. Rather, it is what factor *within the workplace*—the consideration of race or something else (e.g., incompetence, lack of available work, personality conflict)—influenced the decision and produced the adverse event.[69]

---

a worker mangles his arm in a machine—it is presumed to be the "fault" of the enterprise, regardless of the nature of the factors that contributed to it (e.g., worker carelessness). In effect, all causes internal to the workplace are actionable, whereas those which are not lie outside. In most cases, this makes causation easy to determine by simple observation. Likewise, it is easy for employers to monitor potential sources of risk for the actionable events.

67. *See, e.g.*, MICHAEL D. GREEN, BENDECTIN AND BIRTH DEFECTS: THE CHALLENGES OF MASS TOXIC SUBSTANCES LITIGATION 26 (1996); W. KIP VISCUSI, EMPLOYMENT HAZARDS: AN INVESTIGATION OF MARKET PERFORMANCE 264-70 (1979); Richard J. Pierce, *Causation in Government Regulation and Toxic Torts*, 76 WASH. U. L.Q. 1307 (1998); Glen O. Robinson, *Multiple Causation in Tort Law: Reflections on the DES Cases*, 68 VA. L. REV. 713, 721 (1982); David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849, 855-59 (1984); Wendy F. Wagner, *Choosing Ignorance in the Manufacture of Toxic Products*, 82 CORNELL L. REV. 773, 776 (1997).

68. For example, a plaintiff seeking recovery for an occupational disease must address whether exposure to a chemical at work caused his stomach cancer, or whether he would have gotten cancer even if he had not been exposed in the workplace. That question must be answered by recourse to epidemiology, statistical data, and numerical analyses. For a discussion of the difficulties of sorting out issues of causation in toxic torts cases, see references cited *supra* note 67.

69. Even in cases of workplace injury in which causation is difficult to sort out—such as the development of an occupational disease that may be due to toxic chemical exposure on the job—it is relatively easier for the employer to monitor workplace exposure to the risk once the offending agent has been identified. Not only do the innocent potential causes (for example, the background sources of disease risk) remain fairly fixed and outside the employer's control, but it is easy to verify, monitor, and control suspected sources of exposure to the potentially offending substance within the workplace. Thus, even if reducing that exposure might be technically difficult and

In sum, causation is always a central issue in discrimination cases, regardless of the liability rule. Because causation for unconscious disparate treatment claims is quite difficult to determine,[70] the "claims costs," or administrative costs of administering a strict liability rule for unconscious disparate treatment actions will always be substantial.[71] Nevertheless, eliminating the need to make fact-intensive determinations surrounding the standard of care is one highly desirable result of choosing a strict liability rule and argues in favor of that rule in the absence of countervailing considerations. One such consideration is that a strict liability rule might generate a greater volume of claims—and more successful claims[72]—than a negligence standard. Although it would appear that this increased traffic could well outweigh any savings to the system from eliminating the due care inquiry in unconscious bias cases, in reality it will not. It is not just that the due care inquiry will be extremely difficult and expensive to carry out. Rather, as the ensuing discussion makes clear, an efficient level of care against unconscious bias cannot possibly be established given the current state of human knowledge.[73] Strict liability

---

expensive, the result can be effectively monitored.

70. *See infra* text accompanying notes 118-32.

71. *See* LANDES & POSNER, *supra* note 62, at 65. Landes and Posner divide the costs of establishing a liability regime into two types. Information costs are those that relate to the difficulties associated with setting a standard of care. *See id.* Claims costs are incurred in "processing and collecting a legal claim—that is, [in] determining damages, causation, and other issues not involving level of care." *Id.* Information costs would certainly be higher under a negligence rule, but claims costs—setting aside considerations going to the number of claims—will be quite high for both strict liability and negligence.

72. *See* POSNER, *supra* note 62, at 179 ("If most accidents that occur in some activity are unavoidable in an economic sense either by taking greater care or by reducing the amount of the activity, . . . the main effect of switching from negligence to strict liability will be to increase the number of damages claims."). There might be a greater number of successful claims under the strict liability rule if only because a negligence standard requires proof of two elements—failure to exercise due care as well as causation—whereas strict liability requires proof of only one of these (causation).

73. Adopting a negligence rule would require fixing a level of "optimal precautions" against unconscious bias. That level would be the one for which the benefits of taking the precautions (in reducing the risk of actionable harm) outweigh the costs. But, as the discussion of precautions for unconscious bias demonstrates, *see infra* Part II.C.1, the problems of determining the level of "optimal precautions" in this area presents unique conceptual and evidentiary difficulties. First, any methods for tracking a reduction in the incidence of actionable harm—that is, unconscious bias—will be quite unreliable. Second, unlike in many other torts contexts, there is no common sense or common law understanding of what reasonable precautions against cognitive bias might mean. The determination of which precautions are "reasonable" requires an initial investment in the resolution of technical questions that can only be answered by cognitive science. Thus, the "information costs" of setting a standard of care are potentially quite high and the practical difficulties extraordinarily formidable. *See* LANDES & POSNER, *supra* note 62, at 126-31; *infra* notes 115-22. Third, there is no reliable method for assessing the costs of taking precautions against unconscious bias, because there is *no* currently available means of reducing cognitive bias at all. Such expertise is well beyond the reach of psychology in its current state. *See infra* text accompanying notes 118-32. Finally, although negligence theory would mandate the broadest possible inquiry into everything that affects actual risk and the cost of reducing risk in setting the standard of care, in practice such a comprehensive analysis is not actually undertaken and indeed is hardly feasible. For example, the due care inquiry does not customarily include consideration of

almost certainly represents the only reasonable candidate for a workable liability regime directed against unconscious bias. Assuming for the purpose of discussion that strict liability will be superior to negligence in this context, we now turn to a consideration of whether imposing strict liability for unconscious bias will advance the goals of deterrence, effective compensation, and insurance.

C. Deterring Unconscious Disparate Treatment

The theory of strict liability is based on a prediction about behavior: ideally, a rational actor forced to bear the full costs of harms generated by his activities will invest in taking care up to the point where the marginal cost of reducing the harms exceeds the reduction in the expected liability payments for the harms.[74] The threat of liability thus produces an economic incentive to invest in risk reduction until that investment is no longer cost effective for the party creating the risk.[75] The efficiency of a strict liability rule depends on the compensation the defendant expects to pay accurately reflecting the social costs of the harm produced by the defendant's activity. The model assumes that reductions in harm generated by the defendant's efforts will lead to a proportional reduction in expected liability exposure for that party.[76] Beyond that, deterrence will actually occur only if there are feasible methods for reducing harm, and those methods are cost-effective—that is, the value of harm

---

activity levels, although arguably it should. *See, e.g.*, Howard Latin, *Activity Levels, Due Care, and Selective Realism in Economic Analysis of Tort Law*, 39 RUTGERS L. REV. 487, 489-90 (1987). Nor does it always look at the costs of eliminating sheer inattention or human carelessness. *See* Mark P. Grady, *Why Are People Negligent? Technology, Nondurable Precautions, and the Medical Malpractice Explosion*, 82 NW. U. L. REV. 293, 303-07 (1988).

In the same vein, negligence is a poor vehicle for estimating the projected costs of scientific innovations that might be developed to effect risk reduction in the long term. *But see* discussion of innovation *infra* Part II.C.5.e. Because courts are neither prescient nor scientifically sophisticated, they are especially prone to error in setting standards of care where the creation of risk-reduction methods depends on the development of nascent technologies. Where the development of the science necessary for risk reduction is in its infancy—as it is for unconscious bias—optimal care calculations cannot feasibly take the costs of such innovation into account.

74. *See, e.g.*, Polinsky & Shavell, *supra* note 8, at 878-83 (stating the theory behind internalizing the costs of harms to the risk creator through strict liability); *see also* LANDES & POSNER, *supra* note 62, at 64; SHAVELL, *supra* note 36, at 23; Jennifer Arlen, *The Potentially Perverse Effects of Corporate Criminal Liability*, 23 J. LEGAL STUD. 833, 834 (1994). Provided no other party can affect the risk of the actionable harm (which, as discussed *infra* Part II.C.7, is a questionable assumption in the discrimination context), either a strict liability or negligence rule will induce optimal investments in care, so long as due care is set as the level of care that will maximize net benefits over costs. With negligence, the potential tortfeasor is liable only if he fails to take due care, and therefore will invest only enough to comply with the standard of care. The costs of all other accidents will fall on the victim. But that arrangement is efficient because greater investment in care would reduce risk only at excessive cost.

75. For example, if a $100 investment in precautions will reduce a party's expected damages by $150, the party will make the investment. If liability will be reduced by only $90, the party will not make the investment.

76. *See, e.g.*, Polinsky & Shavell, *supra* note 8, at 878-83. Expected damages are a function of the overall probability of being held liable for a harm multiplied by the value of the relief awarded when liability is found. *See id.* at 874.

reduction exceeds the cost of precautions. At one extreme, cost effective precautions may virtually eliminate the risk. At the other extreme, risk reduction may be very hard to achieve. If cost-effective precautions reduce harm very little, then strict liability will not necessarily change the level of harm very much or at all.

The first question to consider is whether effective deterrence is a reasonable prospect for unconscious disparate treatment. Will imposing strict liability on employers in this context tend to reduce the risk that employees will fall victim to unconscious bias? The answer depends on whether employers will be induced to take effective precautions against the offending behavior. But how would employers take precautions against unconscious bias? Assuming that strict liability as applied to unconscious bias will preserve the preexisting structure of vicarious liability structure for agents' discrimination,[77] then either employers must find a way to reorder the personnel system to minimize the danger of biased decisionmaking among direct line supervisors, or supervisors must learn to minimize their own tendencies in this direction and employers must effectively conscript them to this task. Unfortunately, the state of knowledge in cognitive psychology provides reason to doubt that firms could devise effective programs to help supervisors escape the influence of group-based biases. Likewise, direct-line supervisors cannot be expected to control their own unconscious thought processes to avoid discriminatory decisions.

1. Precautions Against Unconscious Bias

Research in cognitive psychology suggests that biases in judgment stemming from categorical generalizations cannot be reliably manipulated or controlled either by the person harboring those biases or by outsiders seeking to redesign the decisionmaking process to reduce such bias. Inherent features of human cognition prevent individuals from detecting or effectively correcting all but the most egregious biases in their own judgments. And science is not close to achieving an understanding of the rules of human psychology that would enable outsiders to manipulate decisionmaking conditions or teach social actors how to control their own thought processes to reduce or eliminate categorical biases on a systematic basis.

Two recently published review articles, which summarize and analyze a voluminous literature in cognitive and social psychology, make the case against the controllability of "mental contamination"—that is, the unwanted influence of unconscious group-based stereotypes on discretionary social judgments.[78] After surveying the major experimental evidence on bias control, John Bargh asserts that "the evidence of controllability is weaker and more problematic than we would like to believe."[79] He suggests that an optimistic stance towards the controllability of stereotyping is based on an "overestimat[ion of] the degree to which automatically activated stereotypes can be controlled through good intentions and effortful

---

77. For discussion of enterprise liability, see *infra* Part II.C.4.

78. *See* Wilson & Brekke, *supra* note 11, at 119-28; Bargh, *supra* note 11, at 366-78. For discussion and definition of "mental contamination," see *supra* Part II.A.

79. Bargh, *supra* note 11, at 361.

thought."[80] Bargh concurs[81] with the analysis offered by the authors of the other review, Tim Wilson and Nancy Brekke, who conclude that "due to lack of awareness of mental processes, the limitations of mental control, and the difficulty of detecting bias, it is often very difficult to avoid or undo mental contamination."[82] Wilson and Brekke assert that in order for contamination of judgments to be avoided, four conditions must be met. First, the decisionmaker "must be aware of the unwanted mental process." That awareness could conceivably stem from "direct introspective access to the process," or from some type of external evidence indicating that the bias is operating.[83] Second, a person must be motivated to correct the error. Third, even if motivated, the decisionmaker "must be aware of the direction and magnitude of the bias."[84] Finally, the person must have "sufficient control over [the responses] to be able to correct the unwanted mental processing."[85]

The authors explain that in most cases of unconscious mental contamination—and especially where persons are engaging in the type of multifactorial subjective assessment that is common to workplace settings and is the central concern of this Article—these conditions cannot generally be met. First, individuals cannot detect unconscious mental contamination at work, because they have no introspective access to their unconscious processes. As Wilson and Brekke state, "When [persons] form an evaluation of someone, what they experience subjectively is usually the final product (e.g., 'This guy is pretty attractive'), not the mental processes that produced this product."[86] Observations of the "final product" or the outcome of a process of assessment are usually of "little help to the lay person trying to untangle what influenced his or her judgments in everyday life."[87] When it comes to unconscious bias "one is never one's own control group."[88] Trying to detect unconscious bias from the "outside" by looking at the results of a handful of

---

80. *Id.* at 362.

81. *Id.* at 370-71.

82. Wilson & Brekke, *supra* note 11, at 117.

83. *Id.* at 119.

84. *Id.* at 120.

85. *Id.*

86. *Id.* at 121 (parenthetical in original).

87. *Id.* at 122. On this point, see, for example, Bodenhausen & Macrae, *supra* note 15, at 22, 37; Alan J. Lambert et al., *Rethinking Some Assumptions About Stereotype Inhibition: Do We Need to Correct Our Theories About Correction?*, *in* 11 ADVANCES IN SOCIAL COGNITION, *supra* note 15, at 141; Yaacov Trope & Akiva Liberman, *Social Hypothesis Testing: Cognitive and Motivational Mechanisms*, *in* SOCIAL PSYCHOLOGY: HANDBOOK OF BASIC PRINCIPLES 239, 265 (E. Tory Higgins & Arie W. Kruglanski eds., 1996) (noting that "evidential biases may also contribute to 'wishful thinking,' or unwarranted high confidence in desirable hypotheses"); Krieger I, *supra* note 1, at 1285-91; Diederik A. Stapel et al., *The Smell of Bias: What Instigates Correction Processes in Social Judgments?*, 24 J. PERSONALITY & SOC. PSYCHOL. 797, 797 (1998) ("[T]here is no phenomenal experience that reliably accompanies the making of a biased judgment as compared to an accurate judgment. People can feel just as confident about their bad judgments as their good judgments.").

88. Susan T. Fiske, *Stereotyping, Prejudice, and Discrimination*, *in* 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY, *supra* note 16, at 357, 384; *see also id.* at 384-91 (stressing the difficulties in detecting and accurately correcting unconscious biases); Wegner & Bargh, *supra* note 16, at 469-78 (same); Stapel et al., *supra* note 87, at 805-06.

workplace decisions is thus unlikely to yield reliable evidence of mental contamination. Only repeated experimental trials that observe the fate of large numbers of employees under controlled conditions can claim to reveal the influence of unwanted biases on workplace judgments.[89] But this "experimental method is of little help to the layperson who is trying to determine the extent to which a particular judgment is biased."[90]

It is also virtually impossible to identify and correct bias from the "inside"—that is, through introspective processes. Decisionmakers are generally unaware of the magnitude and direction of their own automatic biases. Even if they could willfully activate mechanisms to control and correct for presumed biases, they would have difficulty calibrating the corrective measures because they cannot gauge the precise extent to which particular biases are distorting their mental processes.[91] The decisionmaker's incapacity is grounded in "source confusion"—defined as "the inability to recognize the exact contribution of all the influences on our judgments."[92] Our social responses in everyday life are usually multidetermined: "It is as if our minds were an inscrutable cauldron of mental activity."[93] For example, our evaluation of a job candidate ordinarily "is based on more than one of the candidate's many attributes."[94] A supervisor who judges a worker favorably does not really know how much weight he has placed on speed, accuracy, sociability, responsiveness, aggressiveness, or other attributes. The inability to tease apart different sources of influence on judgment stems from the hidden and inaccessible nature of the mental processes at work, which take place automatically and outside of our awareness. The problem of source confusion is made even worse by the self-confirming nature of stereotypical thinking, which appears to alter the ways in which social information is noticed, remembered, interpreted, and processed at the outset.[95]

---

89. *See* Wilson & Brekke, *supra* note 11, at 121.

90. *Id.* at 122.

91. *See, e.g.*, Bodenhausen & Macrae, *supra* note 15, at 37 (It is difficult to estimate "the degree of bias already present in one's private impressions." This difficulty impedes an understanding of "the degree of correction that is required in order to compensate adequately. Given the fuzziness of impressional metrics, it may be hard to know exactly how much adjustment is required."); *see also* Stapel et al., *supra* note 87, at 803 (noting that subjects can respond to being told that their judgments are biased and need correction, but cannot calibrate their responses).

92. Wilson & Brekke, *supra* note 11, at 129.

93. *Id.*; *see also* Fritz Strack & Bettina Hannover, *Awareness of Influence as a Precondition for Implementing Correctional Goals*, *in* THE PSYCHOLOGY OF ACTION: LINKING COGNITION AND MOTIVATION TO BEHAVIOR 579, 579-86 (Peter M. Gollwitzer & John A. Bargh eds., 1996).

94. Wilson & Brekke, *supra* note 11, at 129.

95. Some research suggests that mental categories sometimes guide subsequent information processing and lead to selective attention and learning that tends to confirm and strengthen expectations established by preexisting stereotypes. Because the subject is unaware that information is being ignored or discounted, he has little reason to question the validity of his assessments. *See, e.g.*, David L. Hamilton et al., *Social Cognition and the Study of Stereotyping*, *in* SOCIAL COGNITION: IMPACT ON SOCIAL PSYCHOLOGY 315 (Patricia G. Devine et al. eds., 1994).

> [S]tereotypes tend to bias information processing in ways that maintain and preserve the existing belief system. People tend to seek and remember information that confirms their stereotypes, and hence at the level of the perceiver's subjective experience, those stereotypes are validated by experience. Thus, the use of a stereotype serves to reinforce its apparent usefulness.

Because information that defies entrenched categories may be selectively discounted or ignored, stereotype-contaminated judgments may tend to reinforce themselves over time.

Given these features of mental contamination, it should come as no surprise that no known interventions can reliably reduce or eliminate dependence on categorical thinking in real-life social encounters.[96] In general, research directed at modifying stereotypes or diminishing their influence is riven by rival theories and contradictory results. The literature reveals that psychologists have taken a range of approaches in attempting to reduce categorical thinking triggered by group membership. These include: enhancing the "salience" of protected traits, imposing decisionmaker accountability, manipulating amount and type of informational inputs, developing criteria to screen decisionmakers for biased thinking, and inducing intensive self-monitoring.[97] On careful inspection, none of these methods offers much promise or provides consistent guidance for workplace actions.

Legal commentators have suggested that enhancing the "salience" of race or sex—by causing a decisionmaker to notice or think about a target's group identity—can create conditions that depress the tendency to stereotype.[98]

Id.; *see also* Hamilton & Sherman, *supra* note 15, at 55 ("From the cognitive perspective, the mechanisms that promote the use of stereotypes can undermine the effectiveness of efforts to change stereotypical beliefs."); *id.* at 48 ("The perceiver 'sees' a pattern of information that seems to provide evidence for the 'validity' of the beliefs that themselves influence the way the information is processed.").

96. In reviewing the literature on "prejudice reduction," Wilson and Brekke state that "in general, these studies find a wide range of seemingly contradictory effects" from interventions. Some have no effect while others trigger inadequate or exaggerated responses in varying directions. Wilson & Brekke, *supra* note 11, at 129-34 (reviewing studies); *see also* Hamilton et al., *supra* note 95, at 314-15 ("Despite several recent efforts to pinpoint the factors underlying stereotype change, there still are no good answers."); Hamilton & Sherman, *supra* note 15, at 47 (reviewing the field, and stating that "despite the importance of this question, the problem of changing stereotypes remains a very real and unsolved dilemma"); Krieger II, *supra* note 1, at 1247 (concluding, after reviewing the literature on bias reduction, that "[w]e need additional theoretical and perhaps even empirical investigations into how to reduce cognitive sources of bias"); Allen, *supra* note 1, at 1323-24 (stating that "[t]here is a surprising lack of available information about how to reduce any sort of racism . . . [since] there is little published information on the effectiveness of any given technique"); Bargh, *supra* note 11, at 361-66 (casting doubt on the evidence that stereotyping can be controlled).

There are, of course, two reliable methods for eliminating race or sex-based biases in social evaluation and in occupational settings in particular. One is what Wilson and Brekke term "exposure control": depriving the discretionary decisionmaker of *any* information or knowledge about the target's protected characteristic. Wilson & Brekke, *supra* note 11, at 134-36. Blind law school grading is one such mechanism. Orchestra auditions behind a curtain is another. *See, e.g.*, CLAUDIA GOLDIN & CECILIA ROUSE, ORCHESTRATING IMPARTIALITY: THE IMPACT OF "BLIND" AUDITIONS ON FEMALE MUSICIANS (National Bureau of Econ. Research Working Paper No. 5903, 1997). Alternatively, the subjective element in decisionmaking can be completely eliminated by shifting to exclusive reliance on an objective process, such as multiple choice ability testing, that leaves no room for the decisionmaker's discretion. *See, e.g.*, Kelman, *supra* note 30, at 1158; *see also supra* Part I.A (noting that these options are not feasible in most industrial settings).

97. For general reviews, see Fiske, *supra* note 88; Wegner & Bargh, *supra* note 16; Wilson & Brekke, *supra* note 11; Bargh, *supra* note 11.

98. *See, e.g.*, ARMOUR, *supra* note 1, at 139-40; Johnson, *supra* note 1, at 1032-33.

Alternatively, stressing common interests across groups, rather than enhancing awareness of group identity, may help control stereotypical habits of thought. For example, the contact hypothesis posits that working conditions that encourage individuals from different groups to "conceive of themselves as a single, superordinate group" will reduce categorical judgments about persons from other groups.[99] A close look at the experimental evidence, however, reveals that the effects of attempting to manipulate the salience of group membership are equivocal and depend upon specifics of experimental design and approach.[100] A recent comprehensive review of the intergroup contact hypothesis, for example, concludes that attempts to create conditions that encourage solidarity across conventional group boundaries will sometimes appear to reduce bias and at other times

---

99. For a recent review of the contact hypothesis see, for example, John F. Dovidio et al., *Intergroup Bias: Status, Differentiation, and a Common In-Group Identity*, 75 J. PERSONALITY & SOC. PSYCHOL. 109, 109 (1998).

100. Psychologists have tried to make women or minorities more noticeable by manipulating the ratio of persons of different sexes or races in a social setting or by changing the order of presentation of persons or information to be evaluated. *See* Hamilton & Trolier, *supra* note 15, at 157-58; Shelley E. Taylor, *A Categorization Approach to Stereotyping*, *in* COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR, *supra* note 24, at 83, 89-100.

For example, subjects' responses to "target" persons of a particular race or sex are sometimes observed to depend on whether the targets are well-represented within a group or are in the clear minority. In the latter case, persons are said to function as "tokens" or "solos" in a social setting. In one experiment, token individuals observed interacting at a meeting were judged "as more active and talkative during the discussion, as having had more influence on the group discussion, and [were] rated more extremely on personality characteristics." Hamilton & Trolier, *supra* note 15, at 135. Experimental subjects also tended to recall more of what tokens said in the course of a group interaction. *See id.*

Other researchers report complex frequency effects in which subjects' evaluations of persons within a group depend on the ratio of persons of different races present. *See* David L. Hamilton et al., *The Formation of Stereotypic Beliefs: Further Evidence for Distinctiveness-Based Illusory Correlations*, 48 J. PERSONALITY & SOC. PSYCHOL. 5, 14-16 (1985); David L. Hamilton & Robert K. Gifford, *Illusory Correlation in Interpersonal Perception: A Cognitive Basis of Stereotypic Judgments*, 12 J. EXPERIMENTAL & SOC. PSYCHOL. 392, 405 (1976); Hamilton & Trolier, *supra* note 15, at 136. Yet other studies show that, depending on the behaviors or interactions being witnessed, subjects generally have a more extreme positive or negative reaction to "solo blacks" than to black (or white) persons in balanced groups. *See* Taylor, *supra*, at 89-100; *see also* John B. McConahay, *Modern Racism, Ambivalence, and the Modern Racism Scale*, *in* PREJUDICE, DISCRIMINATION, AND RACISM, *supra* note 15, at 91, 102-10 (noting that black job applicants were judged more negatively or more positively than white applicants, depending on the order in which résumés were presented and whether decisionmaker rates "high-prejudice" or "low-prejudice" on responses to a questionnaire). Other experiments use "priming" techniques that subliminally attempt to trigger stereotypical categories by exposing subjects to suggestive words or stimuli. *See, e.g.*, Bodenhausen & Macrae, *supra* note 15, at 22-34 (Priming of stereotypic thoughts can lead to initial suppression but later "rebound" effects that make categories harder to forget.); Blair & Banaji, *supra* note 14, at 1143. The effects of the "salience-enhancing" devices observed in all these studies are often hard to characterize as positive or negative in their influence on the "harmful" use of group-based biases. Rather, the results of the experiments are distinctly mixed and are dependent on context and experimental design in detailed and often idiosyncratic ways that generate few usable lessons for the range of evaluative situations encountered in the workplace.

exacerbate it.[101] In general, the implications of the work on salience and group mixing for real-life contexts is ambiguous.

Psychologists have also examined the effects of introducing devices designed to enhance experimental subjects' "accountability." Subjects are asked to explain the reasons for their evaluative decisions, or are told they will be rewarded for making assessments that are judged "accurate" or "correct."[102] The results of this research have also been ambiguous and have generated few useful lessons for the employment setting.[103] Others have looked at the effects of exposing the

---

101. *See* Dovidio et al., *supra* note 99, at 110.

102. *See, e.g.*, Bargh, *supra* note 11, at 371-72 (discussing work on "accuracy motivation" in controlling stereotypes).

103. *See* Ralph Erber & Susan T. Fiske, *Outcome Dependency and Attention to Inconsistent Information*, 47 J. PERSONALITY & SOC. PSYCHOL. 709, 724-25 (1984); Susan T. Fiske, *Controlling Other People: The Impact of Power on Stereotyping*, 48 AM. PSYCHOLOGIST 621, 627 (1993); Randall A. Gordon et al., *The Effect of Applicant Age, Job Level, and Accountability on Perceptions of Female Job Applicants*, 123 J. PSYCHOL. 59, 61-66 (1988); Steven L. Neuberg & Susan T. Fiske, *Motivational Influences on Impression Formation: Outcome Dependency, Accuracy-Driven Attention, and Individuating Processes*, 53 J. PERSONALITY & SOC. PSYCHOL. 431, 434-41 (1987); Philip E. Tetlock, *Accountability: A Social Check on the Fundamental Attribution Error*, 48 SOC. PSYCHOL. Q. 227, 230-33 (1985); Tetlock, *supra* note 33, at 76-80; Philip E. Tetlock & Jae Il Kim, *Accountability and Judgment Processes in a Personality Prediction Task*, 52 J. PERSONALITY & SOC. PSYCHOL. 700, 702-06 (1987); Thomas E. Nelson et al., *Everyday Base Rates (Sex Stereotypes): Potent and Resilient*, 59 J. PERSONALITY & SOC. PSYCHOL. 664, 666-69 (1990).

One example of making subjects "accountable" is to offer a monetary reward for successfully cooperating in a task with a person the subject was previously required to evaluate. *See* Fiske, *supra*, at 627; Neuberg & Fiske, *supra*, at 34-35. Alternatively, subjects are asked to justify or explain their evaluations of job applicants to a person in authority. *See* Gordon et al., *supra*, at 61-64; Tetlock & Kim, *supra*, at 702-04; Tetlock, *supra* note 33, at 76-80. According to one of the leading researchers in the field, some of these studies suggest that accountability "can, under certain conditions, motivate people to become more vigilant and self-critical information processors." Tetlock & Kim, *supra*, at 706. However, it is not clear from the studies discussed whether the observations of increased "vigilance" correspond to any actual reduction in reliance on race or sex-based stereotypical assumptions.

Some studies use longer processing times, for example, as a marker for "individuated" rather than "category-based" processing. *See* Neuberg & Fiske, *supra*, at 434. The relationship between long processing times and prejudice reduction is unexplored. Many of the studies suffer from a similar absence of any clear tie between the procedural parameters of "vigilance" or "self-criticism" and more "accurate" or less biased assessment, and none purport to establish a benchmark for measuring whether "accountability" interventions result in a better or more "correct" evaluation of target persons. Nor do they incorporate any direct measure of changes in reliance on race or sex-based mental categories. Moreover, the results of these studies are erratic, suggesting that "accountability does not always lead to greater cognitive work." Tetlock, *supra* note 33, at 75; *see, e.g.*, Nelson et al., *supra*, at 671 (observing futility of attempts to induce disregard of base rates for male and female height through monetary rewards); Nelson et al., *supra* note 15, at 23-30 (reporting a number of experiments in which "a variety of motivational tactics" failed to "destereotype" judgments). Indeed, there is evidence that imposing accountability conditions can sometimes cause subjects to take positions that they believe will please those to whom they are accountable. *See* Tetlock, *supra* note 33, at 80-82 (1983) (reviewing findings that suggest that "accountability motivates cognitive work only when subjects do not have the lazy option of expressing views that they are confident will gain the approval of the person to whom they feel

decisionmaker to greater amounts or a different mix of information about the target. Although some studies suggest that decisionmakers rarely ignore individuating information,[104] there is no good evidence that simply exposing the decisionmaker to more detailed information about a target can be relied upon to reduce the influence of categorical assumptions.[105]

---

accountable"). Finally, in some cases, efforts to impose negative consequences on subjects for "wrong answers" or inaccurate assessments actually lead to "more stereotypical impressions of all applicants." Gordon et al., *supra*, at 59; *see* Wegner & Bargh, *supra* note 16, at 473-74 (describing how accountability and attention to stereotyping can backfire or produce "ironic effects" of heightened reliance on stereotypes); Fiske, *supra* note 88, at 390. For extensive reviews of studies of the effects on judgment of manipulating personal stake or accountability conditions, see Nelson et al., *supra* note 15, at 30-32; Trope & Liberman, *supra* note 87, at 254-58; *see also id.* at 265 (concluding that "accuracy motivation does not necessarily eliminate confirmation biases" and that accuracy incentives "may not prevent implicit evidential biases that favor hypothesis confirmation").

104. *See, e.g.*, Victor Ottati & Yueh-Ting Lee, *Accuracy: A Neglected Component of Stereotype Research*, *in* STEREOTYPE ACCURACY: TOWARD APPRECIATING GROUP DIFFERENCES, *supra* note 23, at 29, 45 ("In fact, perceivers rarely completely disregard individuating information about a social category member."); *see also* McCauley, *Are Stereotypes Exaggerated?*, *supra* note 23, at 215, 238-41 (pointing out that the influence of cognitive categories does not rule out the use of individuating information, if available); Kunda & Thagard, *supra* note 24, at 290-91 (stating that dominance of stereotypes versus individuating information depends on baseline amount of information provided and on particular judgment task).

105. As a general matter, research yields an uncertain answer on whether more or less "individuating" information about targets makes for better or more accurate judgments. *See, e.g.*, Arie W. Kruglanski, *The Psychology of Being "Right": The Problem of Accuracy in Social Perception and Cognition*, 106 PSYCHOL. BULL. 395, 400 (1989) (noting that the "relation between the amount of information and accuracy seems complex" and that "processing more and more seemingly relevant information does not necessarily improve one's chances of reaching a correct judgment"). There is some limited evidence, however, that providing more individualized information about targets can reduce reliance on stereotypes under some circumstances. *See, e.g.*, Nelson et al., *supra* note 15, at 30-36. *But see id.* at 32 (noting that even the most suggestive research "points to continuing category effects, even under conditions that favor individuation"). *See also, e.g.*, Swim et al., *supra* note 14, at 421 (summarizing multiple studies suggesting that, in general, "women will be rated less favorably than men when less information is presented," and that "the amount of information provided may influence effect size"). The studies reported by Swim and her co-authors generally compare the effect of providing little or no information (except sex) about a target with providing detailed individuating information. *Id.* at 422 (citing Anne Locksley et al., *Sex Stereotypes and Social Judgment*, 39 J. PERSONALITY & SOC. PSYCHOL. 821 (1980)). Clearly this comparison does not mirror the workplace, where employers possess, as a baseline, a fair amount of information in addition to group identity about both current and prospective employees. The studies described above can say little about whether providing even more information on top of substantial amounts already available would have any stereotype-abating effects.

There is evidence from some studies that the degree of stereotyping may vary with the qualities and attributes reflected in the information provided—including, for example, whether the information is stereotype consistent or inconsistent. *See* Fiske, *supra* note 88, at 385-86; Lee J. Jussim et al., *Why Study Stereotype Accuracy and Inaccuracy?*, *in* STEREOTYPE ACCURACY: TOWARD APPRECIATING GROUP DIFFERENCES, *supra* note 23, at 3, 13. This suggests that targets who acquire outstanding qualifications or otherwise possess attributes that are atypical of groups otherwise judged negatively might perhaps be assessed in less stereotypical ways. See *infra* text

Alternatively, researchers have looked at whether exposure to information that tends to disconfirm a stereotype might "affect the preexisting stereotypic beliefs."[106] This work has generated a number of conceptual hypotheses concerning whether, and how, exposure to "new" information can alter stereotypical thinking. But none has emerged as the clear winner or produced a workable program for action.[107] Attempts to screen out overt bigots likewise do not offer much promise. Cognitive generalizations and stereotypical thinking can be subliminally triggered in

---

accompanying notes 240-46 for a discussion of "moving targets." The availability of such information in real life depends on variations in the actual profile of the person being assessed, which in turn depends on the target's qualities, attributes, and qualification. Those parameters are sometimes within the control of the person being evaluated. *See infra* note 120. But assuming a target with given attributes (a "fixed" target), it is not possible based on available evidence to say that more information about that target rather than less always serves the goal of reducing reliance on group-based biases. And the literature provides very little guidance on when more information does serve this purpose.

106. Hamilton & Sherman, *supra* note 15, at 49; *see also supra* note 105.

107. According to the "bookkeeping" model of stereotype change, categories are continuously and gradually adjusted in light of information and experience that "either strengthen[s] or weaken[s] existing stereotypes by some modest amount." Thus, "stereotypes are constantly in a state of on-line revision as new information is incorporated." Hamilton & Sherman, *supra* note 15, at 50; *see also* Myron Rothbart, *Memory Processes and Social Belief, in* COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR, *supra* note 24, at 145, 178. The "conversion model" posits sudden and dramatic paradigm shifts in response to the accumulation of a sufficient amount of stereotype-disconfirming evidence. *See* Hamilton & Sherman, *supra* note 15, at 50; Rothbart, *supra*, at 176. Finally, the "subtyping" hypothesis predicts that disconfirming evidence causes broad categorical schema to be broken down into smaller subcategories. The subcategories do not abolish the influence of preexisting beliefs and assumptions, but lead them to be "applied less generally." Hamilton & Sherman, *supra* note 15, at 50; *see* Marilynn B. Brewer et al., *Perceptions of the Elderly: Stereotypes as Prototypes*, 41 J. PERSONALITY & SOC. PSYCHOL. 656, 668 (1981).

Evidence for any of these models is inconclusive. For example, work on subtyping reveals that disconfirming information–as with exposure to a person who clearly defies a stereotype–is often quite ineffective in changing entrenched categorical expectations because people will simply distinguish the exceptional person from the group. Alternatively, they will classify that person within an ancillary subgroup or category that leaves larger cognitive categories essentially intact. Hamilton and Sherman give the example of

> the football player who earns straight As, prefers fine wine, classical music, and English literature to beer, hard rock, and comic books, and shows warmth and sensitivity in his interpersonal relationships. [That person] will not be categorized as a "jock," but rather as a "sophisticate" who happens to play a sport. That is, because he so completely violates the "jock" stereotype, he is not perceived as a member of that category.

Hamilton & Sherman, *supra* note 15, at 53.

Hamilton and Sherman summarize the evidence in this area as suggesting that "there is a delicate balance between the extent to which an individual provides stereotype-disconfirming information and the ability of that information to have an impact on the preexisting stereotype." *Id.* at 53. They conclude that "presenting examples of individuals who strongly disconfirm stereotypic expectancies can be expected to have little, if any, effect, because those individuals will not be viewed as category members." *Id.* at 54. Rather, a more effective vehicle of change might be an individual who "provide[s] some disconfirmation of the stereotype," is perceived "as a 'good' member of the target group," but does not depart too radically from the stereotype. *Id.*

individuals categorized as both "high prejudice" and "low prejudice" based on answers to questionnaires or observations of behavior in experimental situations. Research indicates that persons with tolerant or progressive views of outsider groups are not immune from behavior that suggests subliminal biases.[108]

Finally, some legal commentators have suggested that the adoption of simple or commonsense mental devices, such as engaging in introspective self-criticism or attempting to feel empathy for people who are "different," will go a long way towards banishing cognitive bias from persons' thinking. For example a decisionmaker should routinely force himself to consider the possibility that bias is at work.[109] He should attempt to "individuate" the decisionmaking process by seeking out and relying more heavily on specific information about the person being evaluated.[110] Alternatively, he should try to imagine that he and the employee are the

---

108. *See, e.g.*, Bargh, *supra* note 11, at 364-65 (arguing that activation of subliminal stereotypes does not track tendency to express overt prejudice); Timothy D. Wilson et al., *A Model of Dual Attitudes*, PSYCHOL. REV. (forthcoming 1999) (manuscript at 11, on file with author) (stating that "a number of studies have found low correspondence between implicit and explicit measures of prejudice"); Biernat & Manis, *supra* note 24, at 18 (noting the direction of stereotyping is unpredictable and incidence erratic); Margo J. Monteith, *Self-Regulation of Prejudiced Responses: Implications for Progress in Prejudice-Reduction Efforts*, 65 J. PERSONALITY & SOC. PSYCHOL. 469, 469 (1993) (citing studies which "found that the vast majority of low prejudiced subjects" were "prone to prejudice-related discrepancies"); *see also* Nelson et al., *supra* note 15, at 28-29 (noting studies suggesting that gender role ideology is not predictably related to the tendency to make stereotypical judgments of experimental targets); John Duckitt, *Psychology and Prejudice: A Historical Analysis and Integrative Framework*, 47 AM. PSYCHOLOGIST 1182, 1189 (1997) ("[T]he cognitive paradigm is seriously incomplete . . . . [I]t has made little if any contribution to explaining individual differences in intergroup attitudes and behavior."); McConahay, *supra* note 100, at 99 (finding persons who score high on prejudicial sentiments or attitudes will sometimes discriminate in favor and sometimes against members of minority groups, depending on context and circumstance); Taylor, *supra* note 100, at 100 (demonstrating that subjects vary considerably and somewhat unpredictably in the degree to which they sex-stereotype); *id.* at 110 (use of stereotypes is "remarkably fluid and dependent on the features of the context in which persons are observed"); Wegner & Bargh, *supra* note 16, at 472-73 (stating that stereotypic behavioral response to subliminal racial stimuli is "not moderated by participants' Modern Racism scores").

109. *See* ARMOUR, *supra* note 1, at 139-41 (recommending that persons make an effort to "recall their personal beliefs" when making judgments that might be infected by stereotyping). David Oppenheimer suggests that, when action is taken against a minority job applicant, the decisionmaker should "instantly stop and examine his or her own motives." Oppenheimer, *supra* note 1, at 970. Oppenheimer asserts that if the decision "cannot be justified with a reasonable nondiscriminatory reason," then the decisionmaker should conclude that "the decision may have been negligently reached." *Id.* Oppenheimer provides no further guidance, however, on how the decisionmaker is to determine whether the decision was in fact negligently reached, although aspects of his discussion suggest that perhaps objective validation (that is, the ability to demonstrate a link to some kind of result, like workplace productivity) should be the test. This would, in effect, require that all decisions survive a standard reserved for claims of disparate impact. *See supra* notes 20, 54 (comparing disparate impact to disparate treatment claims).

110. *See* Selmi, *supra* note 1, at 1297-98.

same race.[111] Or he simply should take responsibility for his own intentional acts.[112] The commentators generally assume that these self-help methods will work better if the protected attribute is called to the decisionmaker's attention or is made more "salient."[113] But, as already discussed, attempts to enhance the salience of protected traits produces mixed results.[114] In any event, the studies cited by these commentators offer no convincing evidence that decisionmakers' attempts to manipulate their own mental processes will enable them to identify judgments contaminated by group-based generalizations, will produce well-calibrated reductions in the influence of such biases, or will issue in either more "neutral" or "better" decisionmaking. Some cognitive and social psychologists do appear to express greater optimism than Wilson and Brekke about the prospects for avoiding or correcting stereotypic inclinations in the judgment of specific individuals, but the evidence that subjects can reliably and precisely control stereotyped thinking is remarkably thin. Informing subjects that their conclusions are probably infected with bias, or urging subjects to "try harder" or to "be more careful" has been observed to induce some subjects, for example, to take more time with decisions or to consider a broader range of information. But the great majority of pertinent studies contains no convincing evidence that these shifts in decisionmaking strategy necessarily make for more accurate or "more neutral" subjective appraisals of others. The external criteria or experimental controls necessary for drawing such conclusions are almost always lacking.[115] This point applies with particular force to

---

111. *See, e.g.*, Sheri Lynn Johnson, *Racial Imagery in Criminal Cases*, 67 TUL. L. REV. 1739, 1799-1802 (1993); Johnson, *supra* note 1, at 1032-33.

112. See McGinley, *supra* note 1, at 1470-71, discussing the work of psychologist Susan Fiske. Fiske argues in favor of holding decisionmakers responsible for unconscious bias. She asserts that the decisionmaker "knows how to individuate" in the assessment of a person's attributes and qualifications. She also contends that the process of evaluating an employee is "potentially controllable" because the supervisor could, in some sense, "choose" to evaluate him differently. *See* Fiske, *supra* note 101, at 626; *see also* Susan T. Fiske, *Examining the Role of Intent, in* UNINTENDED THOUGHT 253, 276 (James S. Uleman & John A. Bargh eds., 1989).

Fiske's analysis has been questioned implicitly by other social psychologists and is flawed by fundamental errors in logic and lack of experimental support. *See, e.g.*, Bargh, *supra* note 11, at 365-66 (calling into question conclusions of Fiske and her coworkers). Wilson and Brekke's analysis makes clear that, even if a supervisor chooses to focus more intently on "individuating information" and then determines to alter his judgment, it does not follow that he has succeeded in banishing unconscious trait-based bias from his thought processes. *See* Wilson & Brekke, *supra* note 11, at 130-37. By definition, the true character of the supervisor's decision—the causal antecedents—are hidden from him. That he can "choose" to act differently does not guarantee he will act in an unbiased way.

113. *See, e.g.*, ARMOUR, *supra* note 1, at 139-53.

114. *See* discussion *supra* text accompanying notes 96-103.

115. Jody Armour, a leading "precaution optimist," argues that enhancing the decisionmaker's consciousness of the race or sex of a "target" can help the decisionmaker "consciously monitor [his] habitual responses," and "resist prejudice-like responses when making judgments about . . . group member[s]." ARMOUR, *supra* note 1, at 150. But Armour relies on studies that have little obvious relevance to assessments of employee performance in the workplace. Rather, those studies show that using various devices to call race or gender to experimental subjects' attention will cause the subjects to vary their responses on questionnaires to express fewer overtly stereotypical attitudes when asked to offer a description of themselves or of a target person. *See*

the multifactorial, subjective, "clinical" judgments of the type with which this Article is centrally concerned.[116] Indeed, for the reasons detailed by Wilson and

---

id. at 140. Armour ignores equally strong evidence that "thinking harder" or focusing on trying to avoid bias can backfire by increasing reliance on stereotypes or producing greater errors in judgment. *See* Fiske, *supra* note 88, at 390 (noting how stereotypes may "rebound with redoubled force" when they are consciously controlled); Wegner & Bargh, *supra* note 16, at 473-74 (same); Timothy D. Wilson et al., *Introspecting About Reasons Can Reduce Post-Choice Satisfaction*, 19 PERSONALITY & SOC. PSYCHOL. BULL. 331, 332 (1993) (same); Wilson & Brekke, *supra* note 11, at 127 (same); Timothy D. Wilson & Jonathan W. Schooler, *Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions*, 60 J. PERSONALITY & SOC. PSYCHOL. 181, 191 (1991) (same).

A number of cognitive psychologists, including Patricia Devine, Susan Fiske, and their colleagues, have argued, perhaps in some tension with Wilson and Brekke, that automatic stereotyping can be controlled at times by conscious processes. Upon close inspection, however, their evidence for these conclusions consists primarily of observations about the way in which subjects shift their decisionmaking strategies (e.g., they take more time with decisions, appear to consider more information, and express opinions or offer descriptions of targets that are more often at odds with stereotypic expectations). These researchers do not offer any systematic assessment of the kinds of subjective, evaluative judgments that are the central concern of this Article. What is missing from the studies is a demonstration of the greater accuracy or group-independence of performance assessments that subjects have attempted to "correct" through conscious efforts to control stereotypes. Such evidence would require controlled experimental conditions or an independent criterion of accurate or neutral appraisal. *See, e.g.*, Devine, *supra* note 24; Erber & Fiske, *supra* note 103; Neuberg & Fiske, *supra* note 103. For a critical assessment of Fiske's work on controlling stereotypes, with some confounding findings, see Nelson et al., *supra* note 15, at 15-19, 31. For a critique of Devine's work, see Wilson et al., *supra* note 108 (manuscript at 11) (noting that Devine's work focuses "on the activation of stereotypical *knowledge* about members of other groups and not on the activation of affect or *evaluation*") (both emphases added).

116. There is literature suggesting that subjective appraisals can be made more consistent and reliable by breaking the components of such appraisals down into smaller evaluative units and assigning those units some kind of scalar or quantitative value. *See* William M. Grove & Paul E. Meehl, *Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy*, 2 PSYCHOL. PUB. POL'Y & L. 293, 293-95 (1996) (reviewing the "clinical" versus the "actuarial-statistical" judgment literature). Clinical decisionmaking denotes the application of a loose set of criteria in an impressionistic, subjective, and discretionary manner to form a holistic judgment. There is no good evidence that clinical judgments are improved by the individual decisionmaker's efforts to "think harder," to attend to some criteria with greater care, or to change the relative emphasis on factors in the mix. Indeed, as noted, there is evidence that more intensive introspection leads to less accurate or satisfactory decisions. *See* Grove & Meehl, *supra*, at 313; Wilson et al., *supra* note 115, at 338; Wilson & Schooler, *supra* note 115, at 191.

Some improvement can be achieved, however, by shifting from holistic clinical methods to "actuarial" systems. Actuarial systems do not eliminate the subjective element of judgment, but simply try to make it more systematic. In actuarial systems, judges are asked to assign a specific, albeit subjectively derived, rating in discrete categories to the "target" being assessed. Those ratings are then combined using some kind of fixed quantitative formula. For example, a psychiatrist might be asked to make a prediction about the future dangerousness of a mental patient. Ordinarily he would make a clinical judgment based on an overall impression derived from a range of clinical information. But it is possible to devise an "actuarial" method for making the same prediction. A clinician will be asked to rate the patient in discrete categories that are thought to bear on future dangerousness (such as "paranoia" or "oral fixation"). Those numerical ratings will then be combined in some weighted or unweighted manner to yield a precise prediction score.

Brekke and others, it would be surprising if individuals' efforts to control the influence of group-based categories on their judgments proved to be "effective" in the sense of being precisely calibrated to correct the problem. This is not to say that persons lack the power deliberately to alter the outcome of their deliberative processes: persons who come to believe that irrational group-based prejudices are influencing their appraisals of others in a particular direction can choose deliberately to alter or override their bottom-line judgment in a way that they believe corrects for those prejudices.[117] But the appropriateness of the corrective effort assumes there is something to correct. The conclusion that correction is warranted depends, first and foremost, upon knowing which judgments are infected with group-based biases, and how much and in which direction. Because individuals are not privy to the factors that influence their judgments, the conclusion that a particular judgment is in need of correction may or may not be warranted and any compensatory response will almost surely be imprecise and poorly calibrated. As likely as not, the judgment will either be overcorrected or undercorrected. Thus, the wish to be evenhanded and fair in judgment and the power to alter one's own bottom line response will generally not suffice to "neutralize" social judgments. Individual decisionmakers cannot know whether their efforts have succeeded in factoring out the influence of hidden and automatic biases or in moving decisions towards greater neutrality. There is thus no guarantee that, despite deliberate efforts to make a more careful and individuated decision, that result has been achieved.

2. Detecting Unconscious Bias

Because the precise workings of the mind's unconscious mechanisms are hidden from view, a decisionmakers' determination to "be fair" cannot guarantee freedom from unconscious trait-based biases. In addition, research in cognitive psychology has failed to generate a degree of understanding that would permit the reliable and systematic external manipulation of real-world conditions to diminish the influence of trait-based biases in social judgment. But perhaps it would not matter that workplace decisionmakers or their supervisors are unable to eliminate bias if decisions influenced by race and sex could be singled out in other ways. If the results of biased thinking could be readily identified and monitored, that would introduce the possibility of decisionmakers calibrating their responses by a type of

---

The superiority of actuarial systems may be due to the fact that, although ultimately grounded in subjective judgments, they leave less room for inconsistencies or variations in the application of criteria as between cases. *See* Grove & Meehl, *supra*, at 298-99. Grove's and Meehl's observations strongly suggest that actuarial methods for evaluating employees in the workplace might well prove "better" than holistic, subjective judgments. The greater accuracy promised by adopting these methods even suggests that they may be effective in banishing erratic or irrational influences, such as group-based stereotypes, from the evaluation process. Testing that hypothesis, however, requires devising some objective measure or outcome "criterion" that stands apart from the evaluation process itself. Providing such a criterion in the workplace context is quite difficult, since there is no ready method for assessing performance that is independent of the subjective criteria that stand in need of validation. See Arrow, *supra* note 29, at 96, on the difficulties of measuring productivity using objective criteria and instruments.

117. *See* discussion *infra* at Part II.C.5.b (employer alteration of bottom line judgments on members of protected groups as a way to avoid liability for inadvertent discrimination).

feedback control. Biased thinking could then be thought of as a "bad habit" that operates automatically or unthinkingly, but that can be controlled by deliberate attention to results and effort directed at changing those results.[118]

The problem with the "bad habit" analogy is that there is rarely if ever available an external benchmark, lying outside the tainted sphere of subjective assessment, that indicates when a biased decision has been made. There is no ready way to determine the degree to which a racial or gender stereotype is operating to distort judgment in a particular case because there is no readily available baseline measure of how a particular employee would be assessed if the stereotype did not operate. The obvious criterion to be applied in the workplace setting is job performance, but the use of subjective appraisal methods in the first place suggests that objective benchmarks either do not exist or are thought inadequate to capture what is important about performance. The absence of adequate, non-subjective methods for assessing workplace performance makes it almost impossible to assess the claims that subjective evaluations are biased.[119]

Because neither the decisionmaker nor an outside observer can count on knowing whether a person's efforts to "control" unconscious bias are working, it is virtually impossible to construct a system of incentives for the reform of the decisionmaker's "bad habits."[120] The difficulty of developing external criteria of unconscious bias, which makes it hard to distinguish "tainted" decisions from others, is a problem that confronts both employer firms facing liability for discrimination,[121] and a judicial system attempting to administer a liability system that targets unconscious bias. Although employers and courts face somewhat different issues, constraints, and circumstances in trying to identify bias, the challenge of demonstrating the type of unconscious disparate treatment of concern here–that resulting from the good faith application of overtly neutral but largely subjective criteria–is fundamentally the same in both settings.[122] As noted, rarely will conditions observed in the workplace

---

118. *See* ARMOUR, *supra* note 1, at 134-39 (analogizing prejudiced responses to "bad habits" such as nail-biting or smoking, which are driven by unconscious compulsions but can be brought under conscious control).

119. *See, e.g.*, Arvey & Murphy, *supra* note 24 (detailing the difficulties of validating subjective assessment methods); *see also* Grove & Meehl, *supra* note 116.

120. This suggests why Jody Armour's analogy to "bad habits" breaks down. ARMOUR, *supra* note 1, at 134-39. When a person has managed to stop smoking, he immediately knows he has accomplished this goal. Success can be easily verified both by the person himself and by others. This allows a person to adjust his behavior by a continuous process of trial and error. But a supervisor, or his manager, has no similarly reliable way of knowing whether the supervisor has broken the "bad habit" of biased subjective judgment.

121. For a discussion of vicarious enterprise liability for workplace discrimination, see *infra* Part II.C.4.

122. The difficulties employers face are of a different order than those confronting judges. The latter hear evidence tailored to allegations of specific violations centered on discrete conduct, and must depend on parties to supply the information upon which to base their judgments. Although employers are in a better position to monitor the full sweep of workplace practices over the long haul, many are poorly equipped to collect and analyze the data that would suggest subtle discrimination, or consider it a burdensome distraction to do so.

Fundamentally, however, both employers and courts are hobbled by the absence of criteria for good performance untainted by potential biases. Moreover, as Linda Krieger points out, judges and

supply direct evidence of unconscious discrimination, because rarely will there be some objective or "untainted" measure of employee success, desirability, or productivity.[123] In the absence of such benchmarks, decisionmakers are necessarily thrown back on the only evidence that is probative of unconscious motivation: data showing disparities in treatment. Thus claims of unconscious disparate treatment must stand or fall on the ability to compare the fates of different individuals in the workplace. If unconscious bias is subtle or erratic—if it determines outcomes infrequently or intermittently or makes only a small difference overall[124]—the disparities in treatment of otherwise similar employees may be minor and hard to tie to impermissible bias.

Real employees are, of course, rarely comparable: individual employees, the jobs they occupy, and their background risks are often too dissimilar to permit employees to serve as others' controls for determining the risk of unconscious discrimination. Statistical methods like hypothesis testing and regression analysis are often applied to try to correct for factors that may account for differences in treatment. But the effective use of such techniques often requires large cohorts for comparison and analysis. This problem is made worse by the fact that most unconscious bias cases will surely involve not hiring but rather the treatment of current and former employees whose qualifications and on-the-job trajectories will rarely match.[125] In

---

juries, no less than employers themselves, are vulnerable to expectancy-based biases in considering the evidence before them, whether that evidence relates directly to employee performance or to the possibility that a performance was evaluated in a biased manner. *See* Krieger I, *supra* note 1, at 1305.

> Unfortunately, the same biases in causal attribution that commonly distort intergroup perception and judgment . . . can be expected to distort fact-finder determinations as to whether discrimination has occurred in any particular case. We cannot expect the strength or direction of intergroup decision-maker bias to vary significantly from the strength or direction of intergroup fact-finder bias.

Id.

123. *See* Arrow, *supra* note 29, at 96 (suggesting that many issues of proof in discrimination law stem from the fact that "the ability to observe a measure of the individual's marginal productivity . . . do[es] not in general exist").

124. See *supra* text accompanying note 6, Parts I.C, II.C.1, and *infra* p. 1176, on evidence from cognitive research that patterns of expectancy-influenced judgments are context-specific and unpredictable.

125. As Wilson and Brekke point out, the only sure device for "measur[ing] how much people's judgments and inferences are biased" is

> the experimental method, whereby people are randomly assigned to a condition in which they are exposed to a potential contaminant or a control condition in which they are not. By comparing the average judgments of the two groups, researchers can determine whether a certain kind of information has affected people's judgments.

Wilson & Brekke, *supra* note 11, at 121. For a general review of attempts to isolate discrimination as a cause of intergroup economic and occupational disparities, see William A. Darity, Jr., *Intergroup Disparity: Economic Theory and Social Science Evidence*, 64 S. ECON. J. 805 (1998).

One way of overcoming the limitations of "natural experiments" is to carry out prospective social science experiments that seek to isolate or control for key variables. The Urban Institute has conducted "job tester" studies by sending out job applicants of different races. An attempt is made to match the "tester pairs" for age, experience, education, and demeanor. In some cities in which the experiments were conducted, employers were somewhat less willing to hire minority candidates, although the effects were not large. The methodology of these experiments has been

those cases, the potential for statistical techniques to reveal unexplained differences in treatment comes up against the limits inherent in "natural experiments": even in firms employing large numbers of workers, there may be too few workers and decisions of a particular type to control for potentially relevant variables across cases.

The premise behind the demand for statistical analyses in discrimination cases is that differential outcomes for different groups do not alone prove bias. The expectation that a system free of race-based disparate treatment will achieve almost identical outcomes for minority and non-minority employees in rank, remuneration, success, authority, and other rewards and incidents of employment, although indulged to varying degrees as a convention of employment discrimination litigation in the past,[126] is open to serious question in reality.[127] Mere chance alone produces random variations in the operation of systems of selection on target populations.[128] Beyond chance, there is always the possibility of unidentified "systematic

---

criticized. *See* James J. Heckman & Peter Siegelman, *The Urban Institute Audit Studies: Their Methods and Findings, in* CLEAR AND CONVINCING EVIDENCE: MEASUREMENT OF DISCRIMINATION IN AMERICA 187, 212-18 (Michael Fix & Raymond J. Struyk eds., 1993) (describing and criticizing studies); *see also* Brown et al., *supra* note 1, at 1498-1504 (describing studies); Katherine Q. Seelye, *Agents to Go Undercover in Detection of Hiring Bias*, N.Y. TIMES, Dec. 7, 1997, § 1, at 31 (same). As already noted, the usefulness of such devices, or "tester" programs generally, for documenting employer bias (whether conscious or unconscious) is severely limited because bias against existing employees rarely lends itself to controlled examination.

126. *See, e.g.*, International Bhd. of Teamsters v. United States, 431 U.S. 324, 339 n.20 (1977) ("[A]bsent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired."); *see also* Munroe, *supra* note 57, at 227 (discussing the assumption of equal representation absent discrimination as a background convention in employment discrimination litigation); Strauss, *supra* note 29, at 1640 (arguing that productivity differences among groups, if they exist, must be viewed as the transient products of discrimination); *cf., e.g.*, Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond "Damned Lies,"* 68 WASH. L. REV. 477, 477 (1993) (noting and criticizing the presumption of equal representation absent discrimination as a starting point for litigating employment discrimination claims); Douglas Laycock, *Statistical Proof and Theories of Discrimination*, LAW & CONTEMP. PROBS., Autumn 1986, at 97, 98 (noting and criticizing the presumption of equal representation absent discrimination as fundamentally inconsistent with disparate impact theory).

The assumption that all groups are equally qualified and productive, or would quickly become so absent discrimination, also serves as the basis for some defenses of affirmative action. *See, e.g.*, McGinley, *supra* note 1, at 1472 ("[O]ne could argue that employers should be strictly liable for their failure to hire, maintain, and promote a number of members of the protected classes proportional to their presence in the labor pool."); Selmi, *supra* note 1, at 1276 (defending affirmative action by arguing that there is no evidence of skill or productivity differences between groups that would justify their different representation in the workplace); Strauss, *supra* note 29, at 1654-57 (defending numerical standards for hiring).

127. *See* Arrow, *supra* note 29, at 96; William A. Darity, Jr. & Patrick L. Mason, *Evidence of Discrimination in Employment*, J. ECON. PERSP., Spring 1998, at 63, 82; James J. Heckman, *Detecting Discrimination*, J. ECON. PERSP., Spring 1998, at 101, 101.

128. *See* DONOHUE, *supra* note 3, at 262-74; Browne, *supra* note 126, at 500; Kingsley R. Browne, *The Strangely Persistent "Transposition Fallacy": Why "Statistically Significant" Evidence of Discrimination May Not Be Significant*, 14 LAB. LAW. 437, 437-38 (1998).

differences between the relevant populations as a whole."[129] For jobs from lowest to highest, it is always an open question whether the persistence of unexplained variations in job placement and success can be attributed to race or sex, to neutral factors that enhance productivity or have little to do with it,[130] or to criteria that simply correlate with protected traits.[131] So-called "hidden variables" that account for disparate representation can be quite subtle and mysterious even to the employer

---

129. Browne, *supra* note 126, at 500; *see also* Laycock, *supra* note 126, at 98-99. *See generally* Arrow, *supra* note 29 (discussing group differences); Darity & Mason, *supra* note 127 (same); Heckman, *supra* note 127 (same).

Many factors going to the background or "baseline" risk of unfavorable workplace outcomes are within the control of employees themselves. *See* Glen O. Robinson, *Probabilistic Causation and Compensation for Tortious Risk*, 14 J. LEGAL STUD. 779, 797 (1985) (discussing the driver-plaintiffs' control of accident risk as complicating statistical proof in the Ford Pinto litigation). The possibility that victims' own behavior can influence their treatment in the workplace, by introducing a variable that is hard to track, not only compounds the difficulties of making comparisons among employees, but also threatens to undermine the efficiency of assigning liability exclusively to the employer. See the discussion of victims *infra* Part II.C.7.

130. Correlations between employee attributes and desirability to the employer are hard to demonstrate because workplace demands vary so much and because, as noted, productivity is notoriously difficult to measure. *See* FAIRNESS IN EMPLOYMENT TESTING *passim* (John A. Hartigan & Alexandra K. Wigdor eds., 1989); Arrow, *supra* note 29, at 96; Linda S. Gottfredson, *Education as a Valid but Fallible Signal of Worker Quality*, 5 RES. SOC. EDUC. & SOCIALIZATION 123, 136 (1985); Kelman, *supra* note 30, at 1204; E. Douglas Williams & Richard H. Sander, *The Prospects for "Putting America to Work" in the Inner City*, 81 GEO. L.J. 2003, 2030 (1993) (commenting on the difficulty of developing good data on individual productivity, especially across different firms). *But see* Barbara Lerner, *Employment Discrimination: Adverse Impact, Validity, and Equality*, 1979 SUP. CT. REV. 17, 32. In any event, the absence of a clear link between workplace rewards and productivity is never dispositive. Employers may harbor irrational biases. Or informational limits or other cognitive distortions may cause the employer to make mistaken or imperfect choices that correlate with protected traits but are not based on these traits. *See infra* Part II.C.6.

131. This difficulty works both ways. If unconscious bias takes the form of "statistical discrimination," *see infra* Part II.D, it may be quite difficult to show that race influenced the process because, by hypothesis, the employees selected are more productive than the ones passed over. *See* Keith N. Hylton & Vincent D. Rougeau, *Lending Discrimination: Economic Theory, Econometric Evidence, and the Community Reinvestment Act*, 85 GEO. L.J. 237, 252 (1996) ("If race is a relatively good proxy for the information the statistical discriminator does not collect, then the more information an empirical researcher collects in order to test for racial discrimination, the less evidence there will be of discrimination . . . ."). Conversely, if an employer uses valid neutral criteria that correlate with race and thus have a racially disparate impact, the employer may mistakenly be judged to have engaged in race-based disparate treatment. *See, e.g.*, PHILLIP MOSS & CHRIS TILLY, WHY BLACK MEN ARE DOING WORSE IN THE LABOR MARKET: A REVIEW OF SUPPLY-SIDE AND DEMAND-SIDE EXPLANATIONS 45 (1991) ("[T]here is no statistical way to distinguish between growing demand for unmeasured skills that are correlated with race, and growing levels of racial discrimination."); Williams & Sander, *supra* note 130, at 2029 (The observation that whites earn higher wages even after controlling for obvious sources of differential productivity is "often taken as strong evidence of employment discrimination, but this interpretation suffers from an important weakness: the real possibility that some of the independent variables measuring worker 'quality' are systematically biased along racial lines."). In sum, in the real world there are many ways in which race can masquerade as neutral criteria and neutral criteria as race.

himself. Or they may represent factors, such as motivation or demeanor, that employers routinely notice but that are difficult to measure systematically after the fact. Finally, "hidden variables" might include elements that actually influence decisions in practice but that have received little play for historical or political reasons.[132]

Perhaps the most important factor impeding the detection of unconscious bias is the nature of the phenomenon itself. As already stated, whether unconscious trait-based bias is an important influence on workplace decisionmaking is unknown and, in the current state of knowledge, unknowable. If research on the psychology of "mental contamination" reveals anything, it suggests that trait-based stereotyping is not a constant or predictable feature of social interactions. Rather, it is an erratic and inconstant feature of human judgment.[133] Even on those occasions when bias distorts judgment, it may do so only a little, with no measurable deprivation of concrete benefits or rewards. Unconscious bias thus may not be "determinative" of a harmful outcome in every case in which it can be said to play a part.[134] Bias that only intermittently infects decisionmaking or only occasionally determines a harmful outcome will be harder to detect than bias that is often "determinative," because

---

132. For example, measured general intelligence and performance on various tests of aptitude may correlate with traits employers value. There is evidence that measures of intelligence and aptitude correlate with job-related performance over a broad range of occupations. A spate of recent papers shows that controlling for scores on tests thought to measure intelligence or aptitude causes a persistent "wage gap" between black and white employees to narrow or disappear. *See, e.g.*, Alexander Cavallo et al., *The Hidden Gender Restriction: The Need for Proper Controls when Testing for Racial Discrimination*, *in* INTELLIGENCE, GENES, AND SUCCESS: SCIENTISTS RESPOND TO THE BELL CURVE 193, 195 (Bernie Derlin et al. eds., 1997); George Farkas et al., *Cognitive Skill, Skill Demands of Jobs, and Earnings Among Young European American, African American, and Mexican American Workers*, 75 SOC. FORCES 913, 935 (1997); Derek A. Neal & William R. Johnson, *The Role of Premarket Factors in Black-White Wage Differences*, 104 J. POL. ECON. 869, 892 (1996); June O'Neill, *The Role of Human Capital in Earnings Differences Between Black and White Men*, 4 J. ECON. PERSP., Fall 1990, at 42; *see also* Darity & Mason, *supra* note 127, at 67-68 (summarizing regression evidence on relationship between aptitude, earnings, and discrimination); Linda S. Gottfredson, *Reconsidering Fairness: A Matter of Social and Ethical Priorities*, 33 J. VOCATIONAL BEHAV. 293, 294 (1988) (discussing evidence for group differences in capability); Heckman, *supra* note 127, at 106-07; John E. Hunter & Frank L. Schmidt, *Intelligence and Job Performance: Economic and Social Implications*, 2 PSYCHOL. PUB. POL'Y & L. 447, 447 (1996) (exploring evidence of a strong relationship between aptitude, intelligence, and job performance); Darity, *supra* note 125, at 809-14. *Cf.* Selmi, *supra* note 1, at 1261-76 (casting aspersions on evidence of observed correlations between test performance, productivity, and race).

133. *See* discussion *supra* at Parts I.C and II.C.1. *See, e.g.*, Krieger I, *supra* note 1, at 1258 ("Intergroup bias increases or decreases in response to contextual, environmental factors which shapes how social actors perceive, judge, and make decisions about members of their own and other reference social groups."); Bargh, *supra* note 11, at 376 (stressing the evidence "that automatic stereotype activation does not occur for everyone" and that "there may be individual differences in whether [a] stereotype is activated" under different circumstances).

134. See discussion of determinative versus nondeterminative discrimination *supra* text accompanying notes 27-28 and *infra* Part II.D.1.

only a small proportion of all decisions involving protected group members will be affected.[135]

Moreover, even when decisional outcomes are affected, the pattern of disparities generated by unconscious bias may make it quite difficult to detect. The *magnitude* of the effect may be quite small for each victim, and bias will often be a minor factor in the decisional mix. For this reason, differential treatment influenced by unconscious factors will rarely appear arbitrary on its face or wholly unwarranted. The claimant may assert that the discipline he received was somewhat *too* severe or extreme, or that his smaller bonus resulted from somewhat *too much* weight being given to his shortcomings and not quite enough to his strengths. In sum, unconscious discrimination may be "subtle" or elusive in the following ways: it may play a part only in some decisions and not in others; where it plays a part, it may make a detectable difference in only some cases; and where it makes a detectable difference, that difference may be small. Furthermore, the differences in treatment that this pattern of bias creates may show little apparent rhyme or reason. Similar individuals may be subject to quite different degrees of unconscious bias due to subtle differences in presentation or context that are difficult to detect after the fact. All these factors will make it hard to demonstrate bias against a particular individual or against a group overall using conventional methods of proof in discrimination cases. The low "signal-to-noise" ratio that results will contribute to the already formidable difficulties of demonstrating unconscious motives using standard methods of statistical proof.

3. Conscious and Unconscious Bias Compared

The previous discussion suggests that no one knows how to selectively eliminate race or sex-based cognitive biases from subjective evaluation processes. Neither workplace decisionmakers nor their supervisors can detect or correct "mental contamination." But it would then seem to follow that incentives aimed at bringing about this result by internalizing the costs of harm to the relevant actors–here employers–will generally be ineffective. Because employers do not know how to respond constructively to such incentives, there is no reason to believe that the risk of unconscious discrimination will be reduced by encouraging employers to take steps against decisionmaking bias. Put bluntly, liability for unconscious discrimination will not deter unconscious discrimination.

This discussion begs the question of how the analysis differs for conscious and unconscious bias. This Section suggests that the potential for deterring

---

135. It is possible to imagine the existence of evidence for "bias in the process" that operates short of making a "but for" difference to the allocation of tangible rewards or benefits. For example, a plaintiff could seek to show that minorities receive lower ratings on a subjective scoring system for promotions, regardless of whether many or most of those individuals would otherwise have attained a score above the cutoff. *See supra* text accompanying notes 11, 27. But in many other cases nondeterminative discrimination will literally vanish without a trace because there is no record of the process leading to the decision, or that process has unfolded entirely inside the decisionmaker's head. *Cf.* Strauss, *supra* note 1, at 958 n.72 (suggesting that nondeterminative discrimination always entails some kind of less favorable treatment, which is detectable as a difference at least in principle).

discrimination by imposing liability is significantly greater for conscious bias because it is far easier for supervisors, and their employer firms, to take effective precautions against deliberate discrimination than against discrimination that is hidden even from the decisionmaker himself.

The analysis begins with the persons on the front lines–those directly responsible for observing and evaluating workers in the first instance. Supervisors who practice conscious as opposed to unconscious discrimination, like those who inflict intentional as opposed to "accidental" harms generally, differ in their awareness of what they are doing and in their level of control over the injuries they inflict.[136] These two parameters are not unrelated. A central element of consciousness is the ability to observe or evaluate one's own thoughts.[137] The cognitive processes that produce "intentional" actions can be monitored effectively by the actor through introspection, which reveals the nature of what he is doing. Harms inflicted "deliberately" or as the outgrowth of self-conscious motives are also generally amenable to a much higher degree of actor's control than harms produced by processes that are not available to introspection.[138] Finally, because there is generally a tight link between act and result for purposeful harms,[139] the choice not to act *from a consciously discriminatory motive* can result in an effective decision to refrain from inflicting the tortious harm. In contrast, where bias is inadvertent, the actor can always choose to act differently (e.g., by changing the outcomes of his decisions), but he cannot generally make an effective choice not to act *out of bias*.[140]

In light of this discussion, it should be obvious that imposing strict liability for deliberate discrimination has at least the potential to produce more effective deterrence than imposing liability for unconscious discrimination. A conscious discriminator can respond to incentives, because he has a firm internal benchmark

---

136. See discussion of intentional and unintentional torts *supra* Part II.B.

137. *See, e.g.*, Wegner & Bargh, *supra* note 16, at 453-54 (discussing the distinction between conscious and unconscious processes in terms of awareness and control); Devine, *supra* note 24, at 6.

138. Whether motives or reasons for action can be said to "cause" an action or its result is the subject of endless philosophical speculation that is beyond the scope of this Article. *See generally* Gudel, *supra* note 50, at 20; *Symposium on Causation in the Law of Torts*, 63 CHI.-KENT L. REV. 397 (1987). It suffices for our purposes to take note of the common sense understanding that reasons in some sense "produce" actions. Moreover, where actions are taken deliberately—for instance, out of a self-conscious desire to do harm—the actor has the power to refrain from taking the action. Therefore, he has the power to avoid doing the harm, and holding him responsible for that harm is generally seen as unproblematic. *See, e.g.*, Krieger II, *supra* note 1, at 1185 (noting the relationship between the assumption that persons are "aware of the reasons why they are about to make, or have made, a particular employment decision" and the ability of those persons to "comply with Title VII's injunction 'not to discriminate'").

Nonetheless, the ability to cease doing what is done "on purpose" may not be present equally in every case, because some persons may act under a pathological compulsion or exhibit weakness of the will. Some bigots may know very well that they are discriminating against blacks or women but may be unable, in some sense, to stop themselves. But leaving those cases of compulsion aside, the purposeful nature of consciously directed action in most cases offers at least some potential for self-control.

139. *See supra* note 37.

140. For a discussion of activity level effects, see discussion *supra* Part II.C.1 and discussion *infra* Part II.C.5.a.

for the presence of offending conduct. He knows when he is discriminating, and when he is not. And he knows how to take precautions against the harm of discrimination: he can stop discriminating. Moreover, the ability of supervisors to stop discriminating "on purpose" through simple self-help keeps process costs low. Responding to liability does not require the acquisition of specialized technical knowledge or the implementation of sweeping workplace reforms. For irrational discrimination, the costs of choosing to be neutral are the loss of whatever psychic benefits bigots enjoy. For statistical discrimination, there may be additional costs from loss of information. But there are no other significant obstacles to implementing precautions. The theory of strict liability predicts that firms will cease deliberately discriminating if the costs imposed by the liability system outweigh whatever benefits the firms or their employees enjoy from discriminatory behavior.

4. Enterprise Liability and Monitoring Bias

So far, the discussion has focused on supervisory personnel directly responsible for evaluating other employees. Those persons have far more control over their conscious biases then their unconscious ones. But one potential objection to this tidy story is that it fails to take into account the rule of vicarious enterprise liability for discrimination in the workplace.[141] Employer firms are routinely held liable for supervisors' acts of intentional disparate treatment, whether inflicted with the knowledge and blessing of the firm or not. But an enterprise does not have full control over its agents' conscious choices, just as a supervisor does not have full control over his own subconscious processes. Discrimination as practiced by line supervisors against company policy is like an accident for the enterprise in that it is unintentional, inadvertent, and unpredictable from the firm's point of view. The danger of liability in that setting gives rise to a risk management problem for the employer regardless of whether the agent is acting unconsciously or with deliberate intent.

These observations beg the question of why the rule that governs a firm's liability for an agent's purposeful discrimination should be any different from the one that governs the liability of an agent for his own unconscious acts (or of a principal for an agent's unconscious acts). Despite ostensible parallels, there are still sound reasons for applying a different liability rule to the enterprise depending on whether the injuries inflicted by its agents are inadvertent or deliberate. The justification is grounded in differences in the nature of the risks the enterprise must manage and the methods available for managing them.

By attempting to internalize all costs of an activity to the firm, enterprise liability creates an incentive for the firm to take efficient precautions against the risk. But where the risk at issue is of discriminatory harms, the employer-enterprise may have

---

141. *See* Alan O. Sykes, *The Boundaries of Vicarious Liability: An Economic Analysis of the Scope of Employment Rule and Related Legal Doctrines*, 101 HARV. L. REV. 563 (1988) [hereinafter Sykes, *The Boundaries of Vicarious Liability*]; Alan O. Sykes, *The Economics of Vicarious Liability*, 93 YALE L.J. 1231 (1984) [hereinafter Sykes, *The Economics of Vicarious Liability*]; J. Hoult Verkerke, *Notice Liability in Employment Discrimination Law*, 81 VA. L. REV. 273, 341 (1995); Rebecca Hanner White, *Vicarious and Personal Liability for Employment Discrimination*, 30 GA. L. REV. 509 (1997).

to implement its risk-reducing strategy by altering the conduct of its "front-line" agents. The firm can attempt to promote greater neutrality in decisionmaking by introducing structural reforms in its personnel system—a result quite difficult to achieve.[142] Alternatively, the enterprise can try to create effective incentives for the agent to take greater care.[143]

Incentives are of little value if the agent is inherently incapable of responding to them. The problem with automatic bias is that it is refractory to incentives of *any* kind. Even if the agent fears the sanctions the employer threatens to impose, and even if he believes there is a significant probability of detection and punishment, his ability to monitor or control his own inadvertent bias is quite limited. This makes it hard, if not impossible, for him to respond precisely and effectively to the employer's threats. Thus even assuming an enterprise can detect instances of discrimination and identify offending supervisors with a high degree of accuracy (a dubious assumption), threatening to punish supervisors will not reliably reduce the risk of biased evaluations or otherwise neutralize the decisionmaking process where disparate treatment results from unconscious forms of bias.

In contrast, as discussed in the previous Section, the supervisor who deliberately discriminates does have the capacity to respond effectively, and at low cost, to incentives the employer creates to suppress the offending behavior. Then the efficacy of imposing strict liability on the enterprise will depend on the methods available to the principal-firm for inducing its employees to refrain from behavior that exposes the firm to liability. Although the ability to create an effective incentive system depends on a number of factors,[144] an important one is the employer's ability to identify and monitor risk-creating behavior.[145]

The previous Section has already discussed the difficulties inherent in trying to detect and monitor unconscious bias. But deliberate bias can be difficult to oversee as well. Persons motivated by self-conscious animus often try to hide or camouflage their motives, and it is often unclear whether an adverse action against an employee was motivated by discrimination or another factor. Since employers cannot look into the minds of their agents any better than unconsciously biased agents can look into their own, firms attempting to monitor deliberate animus, like those trying to detect unconscious discrimination, will be forced to rely on external cues, including statistical analysis of outcomes, with all the problems inherent in those methods.

Despite these parallel difficulties, there are nevertheless good reasons, from the point of view of deterrence, to adopt a liability system that treats unconscious and conscious bias differently. First, purposeful bias does not always implicate the problem of monitoring renegade supervisors because it may take the form of invidious policies or practices maintained at the highest levels. In those instances,

---

142. See discussion *supra* text accompanying notes 11-38 and *infra* text accompanying notes 161-68 of the difficulty of creating personnel structure that is "bias free."

143. *See* Sykes, *The Boundaries of Vicarious Liability*, *supra* note 141, at 1237; Verkerke, *supra* note 141, at 341-44.

144. *See* Sykes, *The Boundaries of Vicarious Liability*, *supra* note 141, at 578-79 (discussing the challenge of creating an effective penalty structure for employees in the teeth of the employee's limited financial resources, financial exposure, and job investment).

145. *See id.* (stressing problems of information and monitoring); Sykes, *The Economics of Vicarious Liability*, *supra* note 141, at 1231 (same); Verkerke, *supra* note 141, at 341-43 (same).

imposing liability on the firm has the potential to deter unlawful practices directly. Second, unconscious bias is, by definition, a problem that arises from the use of discretionary and subjective procedures rather than objective methods of workplace evaluation. The Supreme Court and others have recognized that subjective bias is notoriously difficult to demonstrate to the satisfaction of the fact-finder.[146] That observation applies regardless of whether the bias is inadvertent or quite deliberate. But the fact that some cases of deliberate or animus-based bias will be difficult for employers to monitor or detect—because embedded, perhaps, in subjective assessments that are difficult to compare—does not mean that all cases will be. Conscious bias may also manifest itself in decisions that are supposed to be governed principally by objective criteria. Where those criteria are disregarded or overridden, bias may be relatively easy to monitor and prove. By definition, however, unconscious bias of the type considered here will infect subjective appraisals only. Because a rule against conscious bias will take on some "easy" cases, whereas the rule against unconscious bias will only take on hard ones, the former is worth maintaining even if the latter is not.

Third, even if the focus is on subjective decisionmaking, purposeful discrimination may in general be easier to detect and monitor than unconscious cognitive bias. Purposeful bias has at least the potential to call attention to itself through anecdotal evidence such as telltale conversations, idiosyncratic remarks, or suggestive workplace behavior that confesses the agent's state of mind. Inadvertent distortions in the decisionmaking process will not generally yield this type of particularistic or anecdotal evidence because, by definition, the agent wishes to be, and believes himself to be, "fair."[147] In addition, purposeful bigotry may well generate effects that are more blatant, extreme, and internally consistent in their operation than inadvertent bias. When a supervisor harbors conscious bias based on race or sex, that factor will often dominate the decisionmaking process.[148] Deliberate bias may have harsher effects that will make it easier to measure: The cruder the bias, the fewer the data points needed to demonstrate it and the more likely a firm can detect, or a plaintiff show, discrimination by comparing a few examples and eschewing complex statistical analysis. Thus, firms may in general have an easier time monitoring animus-based conduct. If, for this reason, employees realistically fear detection more often for deliberate than for inadvertent discrimination, liability for the former will have a more significant deterrent effect.

Finally, and perhaps most importantly, the distinction between conscious and unconscious bias implicates the power of the law to change behavior by expressing

---

146. *See* Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 990 (1988); Strauss, *supra* note 29, at 1655-56.

147. Even if anecdotal evidence of prejudice is available, it is doubtful that it should be considered probative: there is no systematic correspondence between "mental contamination," prejudiced views, and overt expressions of bias. *See supra* text accompanying notes 11-17.

148. Mixed motive cases stand as a potential exception: by definition, some nondiscriminatory factor is also influencing the decisionmaking process. But that factor may be clearly inadequate to explain the outcome, or the employer may otherwise make the invidious motive clear. See *supra* text accompanying notes 57-58 on "mixed motive" cases and the type of "direct" or "anecdotal" evidence usually required.

society's disapproval or its conception of right conduct.[149] Laws against discrimination in the workplace make clear that bigotry is unacceptable. Through their effect on social norms of workplace conduct and individuals' conceptions of justice and fairness—which can operate wholly apart from any legal sanctions—antidiscrimination laws may cause individuals to try to suppress their own biased impulses or at least not to act on them.[150] But this effect will only work when discrimination is deliberately inflicted. The individual's desire to reform will issue in a diminution of conscious bias, because the individual can control discrimination inflicted "on purpose." In contrast, good intentions are of little use when bias is inadvertent, since such biases cannot be reliably purged from social judgment by self-regulation alone. There is thus little to be gained in practice by penalizing unconscious bias for the purpose of sending a message that such bias is "wrong."

5. Employer Responses to Liability

As explained in the foregoing Section, firms and their agents will find it virtually impossible selectively to root out unconscious trait-based disparate treatment from the workplace by changing decisionmakers' behavior. This suggests, at first blush, that internalizing costs to employers in the hopes of inducing them to take cost-effective precautions against biased behavior is likely to be futile because it is impossible for them to take any effective precautions at all. Unconscious discrimination would appear to be an "unavoidable accident." The targeted response of purging all decisions of trait-based bias, however, does not exhaust the range of possible responses that an employer might make to the threat of liability for unconscious bias. An assessment of the cost-benefit balance of imposing strict liability for unconscious disparate treatment is not complete without a consideration of other steps, apart from removing the taint of bias itself, that employers might take in response to the threat of liability.

a. Activity Level and Employment Effects

Where effective precautions against a targeted harm cannot be taken because technologically infeasible or prohibitively expensive, an enterprise may be driven to respond in other ways to the threat of liability for those harms. One strategy is to reduce the level of operations generally or to shift away from the vulnerable activity

149. For a recent discussion of the expressive or symbolic value of law, see, *e.g.*, Richard H. McAdams, *The Origin, Development, and Regulation of Norms*, 96 MICH. L. REV., 338, 398-400 (1997) (arguing that "law also expresses normative principles and symbolizes societal values, and these moralizing features may affect behavior"); *see generally* Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943 (1995) (discussing the power of law to change attitudes and behavior apart from the fear of legal penalties).

150. *See, e.g.*, McAdams, *supra* note 149, at 399 (noting that "several theorists have suggested that the expressive function of law works by affecting *norms*") (emphasis in original).

to alternatives.[151] In the case of strict liability for unconscious disparate treatment, a firm might search for ways to reduce the number of persons employed (by, for example, switching from labor to capital-intensive means of production). Or it might try to employ fewer persons presenting a threat of liability by, for example, relocating to areas with a predominantly nonminority workforce. These effects have been noted as possible consequences of employment discrimination laws generally.[152] A full analysis of the potential for effecting such compensatory adjustments will not be undertaken here, except to note that it is often difficult to predict whether and to what extent a particular enterprise will respond to liability for unavoidable accidents by reducing the scope of its activities. But if the response to liability takes the form of fewer jobs for minority, female, or other workers, that consequence would not ordinarily be regarded as a socially desirable outgrowth of strengthening employment discrimination laws, if only because it would tax many of the same people the laws are designed to help.

Alternatively, if the costs of reducing activity levels are too high or if the activity is still cost-effective after absorbing liability payments, the enterprise may simply continue business as usual and may try to "pass through" the costs of liability to employees, managers, customers, stockholders, or owners, or some combination of these.[153] Such pass-throughs might be considered undesirable for two reasons. First, pure transfers or redistributions of resources are generally thought to impose deadweight losses because they generate transaction costs without any compensating risk reduction.[154] Second, an enterprise might try to shift the unavoidable costs of liability to employees in the form of lower wages or fewer benefits. This would

---

151. *See, e.g.*, SHAVELL, *supra* note 36, at 277 (speculating about activity level effects from incentives created in the employment liability context); Verkerke, *supra* note 141, at 341-44 (same).

152. *See, e.g.*, RICHARD A. EPSTEIN, FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS 61-69 (1992); John J. Donohue III, *Advocacy Versus Analysis in Assessing Employment Discrimination Law*, 44 STAN. L. REV. 1583, 1602-03 (1992) [hereinafter Donohue, *Advocacy Versus Analysis*] (reviewing EPSTEIN, FORBIDDEN GROUNDS, *supra*); John J. Donohue III, *Further Thoughts on Employment Discrimination Legislation: A Reply to Judge Posner*, 136 U. PA. L. REV. 523 (1987) [hereinafter Donohue, *A Reply to Judge Posner*]; John J. Donohue III, *Is Title VII Efficient?*, 134 U. PA. L. REV. 1411 (1986) [hereinafter Donohue, *Is Title VII Efficient?*]; Richard A. Posner, *An Economic Analysis of Sex Discrimination Law*, 56 U. CHI. L. REV. 1311 (1989) [hereinafter Posner, *Economic Analysis of Sex Discrimination Law*]; Richard A. Posner, *The Efficiency and the Efficacy of Title VII*, 136 U. PA. L. REV. 513 (1987) [hereinafter Posner, *The Efficiency and the Efficacy of Title VII*].

153. For a discussion of pass-throughs in the products liability literature, see Alan Schwartz, *Proposals for Products Liability Reform: A Theoretical Synthesis*, 97 YALE L.J. 353, 360 (1988) (assuming, for purposes of analysis, 100% pass-through to consumers through higher prices). *But see* Duncan Kennedy, *Distributive and Paternalist Motives in Contract and Tort Law, with Special Reference to Compulsory Terms and Unequal Bargaining Power*, 41 MD. L. REV. 563 (1982) (suggesting that cost-shifting varies considerably in different liability contexts depending on the type of liability, market conditions, market imperfections, information problems, and other factors).

154. *See* John J. Donohue III, *The Law and Economics of Tort Law: The Profound Revolution*, 102 HARV. L. REV. 1047, 1066-67 (1989) (book review) (noting the "typical argument against wealth transfers—that because they are costly to effect and do not expand the size of the pie, but only change its distribution, they actually reduce total social welfare").

mean that all or some workers would effectively end up bearing the costs of the sums collected by plaintiffs as compensation under the liability system. If minorities bear the brunt of the pass-throughs, the at-risk group would effectively insure itself. The fairness or desirability of this result is a complicated question. Apart from distributional concerns, whether such shifts are efficient will depend on whether the transfers end up reducing or increasing social utility overall, which in turn depends on the marginal utility of the intra-personal and interpersonal shifts in resources represented by the transfers. Since it is almost impossible to predict who will end up paying in any given case,[155] very little can be said about whether a liability rule that yields more cost shifting than activity effects will produce a desirable result.

b. Overinvestment in Precautions

The foregoing Section suggests that, to the extent that strict liability may produce reductions in activity levels, that effect might prove harmful to the very groups that are the intended beneficiaries of the liability system. This Section discusses yet another potential employer response that argues against expressly extending liability to unconscious forms of bias: the danger that employers will respond to liability by engaging in inefficiently expensive precautions against the threat of being held liable.

The argument is based on the prediction that employers will be able to adopt strategies that reduce the risk of liability for discrimination, but without necessarily effecting a proportionate reduction in the amount of group-based unconscious bias in their employment practices. As explained below, the danger of socially inefficient "overinvestment" in precautions is produced by the joint operation of two peculiar features, already discussed, that characterize a system of liability for inadvertent bias: the intransigence of unconscious bias to any precautionary program, and the comprehensive control that employers exert over adverse or "injurious" actions against employees, both discriminatory and benign. As noted above, employment discrimination is unique in that liability turns on distinguishing actionable from non-actionable causes of decisions adverse to employee interests, all of which are internal to the workplace. Not only does this feature make causation hard to sort out, but it enables a particular response to the threat of liability. Specifically, it opens the way for employers to deflect a finding of liability by manipulating the treatment of employees to create the illusion that forbidden biases have been purged from its decisionmaking processes.

Overinvestment in response to liability is a danger in many legal regimes, whether based on strict liability or negligence. Overinvestment can occur, for example, when an enterprise is held responsible for harms it does not create or that are not actionable.[156] Errors leading to liability for nontortious harms can arise from imperfections in the judicial process, including faulty or exiguous evidence, fact-

155. *See* Kennedy, *supra* note 153, at 566.

156. *See* Polinsky & Shavell, *supra* note 8, at 873 ("[I]f injurers are made to pay more than for the harm they cause, wasteful precautions may be taken . . . and risky but socially beneficial activities may be undesirably curtailed.").

finder error, or uncertainties in the legal standard to be applied.[157] There is yet another scenario that can produce an inefficient level of "precautionary" investment in response to a liability threat: when investments allow an enterprise to escape liability without effecting a proportional reduction in the harm targeted by the liability system. Such a pattern might arise from imperfections in the administration of the liability system—as when a judicial fact-finder mistakes the appearance of a cure for the targeted harm with an actual cure.

In both of these overinvestment situations, the problem comes down to wasted expenditures: an enterprise invests resources, but harm is not reduced enough to justify the investment. For liability systems to function properly, the resources a tortfeasor invests in precautions in response to an expected amount of liability must actually reduce the harm with which the liability system is concerned, and must do so without generating significant negative externalities. A party will invest the right amount only if that investment reduces that party's expected liability in proportion to a reduction in the social costs of the targeted harm.[158]

But where tortfeasors can take steps to reduce their potential liability costs without effecting a proportionate reduction in the specific harms for which they are being charged, there is no reason to believe that the system will be cost-effective. This point represents an extension of an argument made by Craswell and Calfee in the negligence context. The authors point out that a particular precautionary investment will be worthwhile for the enterprise if the marginal reduction in expected damage payments from the investment exceeds the extra cost of taking the precautions, wholly apart from any reduction in the expected social costs of harm.[159] Craswell's and Calfee's insight is not specific to the negligence context. Nor is it

---

157. On over-investments in precautions due to uncertainty about the legal standard for liability, see John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965, 981 (1984) (discussing how imprecision in determining the standard of care can produce overinvestment in precautions in negligence regimes). *See also* Krieger I, *supra* note 1, at 1247 (citing John E. Calfee & Richard Craswell, *Deterrence and Uncertain Legal Standards*, 2 J.L. ECON. & ORG. 279 (1963), for the proposition that the unpredictability of legal standards for employment discrimination can produce large variations in employer compliance). As the ensuing discussion makes clear, discriminatory actions are not immune from fact-finder errors that can lead to overinvestment.

158. *See, e.g.*, Polinsky & Shavell, *supra* note 8, at 887-89. As David Rosenberg has stated, "[A] strict liability rule allows a firm to reduce its potential liability only by reducing the probability of accidents." Rosenberg, *supra* note 67, at 866. Strict liability will induce the proper investment in precautions because the firm "will cease investing when it can no longer profitably affect the probability [of liability]." *Id.*

159. *See* Calfee & Craswell, *supra* note 157, at 970. The authors argue that if the investment in precautions causes the risk of liability to drop faster than the actual level of social costs generally, some portion of the precautionary investment will be wasted. *See id.* In their case, it is the legal uncertainty of the standard of due care, and the error that it generates, that creates a disparity between the actual reduction in the costs of the targeted harm brought about by the defendant's precautionary investment, and the reduction in the amount of expected damages that the defendant can expect to pay at this level of investment. *See id.* at 974. In our case, it is the fact-finder error generated by the employer's ability to create an impression that unconscious discrimination has been eliminated from the decisionmaking process, when in fact it may not have been—or may have been only to some degree—that generates the potential disparity between expected liability and social costs. *See* discussion *infra* pp. 1186-93.

peculiar to the type of error discussed by the authors that leads to a disparity in expected damages and expected harm (which is error in the application of the standard of due care to the facts). Rather, it comes into play wherever there is the potential for systematic error in applying *any* legal standard to the facts at hand and the tortfeasor can manipulate that error to his benefit.

The tortfeasor is concerned with the amount of liability, not with social costs. Expected liability is the "signal" that informs the tortfeasor of the social costs he is generating, and of the amount he must expend to reduce the social costs of that harm. But if the signal goes awry—because the actor's expected liability drops off faster than the value of the harm his investments are supposed to prevent—the actor's expenditure may fail to increase net social welfare. The potential for an inefficient response is even greater if a defendant's liability-reducing strategies generate negative externalities or entail massive process costs.[160]

If the goal of a liability system is to reduce unconscious trait-based bias in workplace decisionmaking, the danger of overinvestment in precautions may arise if this scenario obtains: an enterprise can take measures that will reduce the likelihood it will be held liable for discrimination, but without making its personnel system significantly more "neutral" or without effectively reducing the influence of unconscious bias on its decisionmaking processes. Could this happen? It could happen if the enterprise could take steps to convince judicial fact-finders that it was not making unconsciously biased decisions or that it had effectively controlled unconscious bias even if those steps reduced unconscious bias erratically or not at all, and even if they generated undesirable social costs. The fact-finder would then underestimate the amount of actionable discrimination in which the defendant was actually engaged, and would reduce penalties accordingly. If the penalty reduction exceeded the cost of the "precautionary" measures, employers would invest in those measures regardless of any effect on the targeted harm.

Although corporate programs aimed at addressing the problems posed by a diverse workforce can take a variety of forms, there are two main types of strategies an enterprise might adopt in response to a threat of liability for unconscious discrimination. The first is to implement "diversity awareness" programs, which are usually short educational or training sessions designed to make people more aware of their prejudices against protected groups and how their biased attitudes can impede the progress of minorities and women. These training programs have become quite widespread.[161] A second strategy is to introduce more far-reaching

---

160. For example, a party will willingly spend $6,000 to avoid $12,000 in damages costs, even if that expenditure generates only $3,000 worth of harm reduction and creates an additional $3,000 in social costs or "negative externalities." In that case, the employer would spend $6,000 to effect no net change in social welfare ($3,000 in targeted harm reduction offset by $3,000 in external costs). A $6,000 expenditure for nothing is socially inefficient. Yet an employer will voluntarily spend $6,000 because it saves $12,000 in liability costs.

161. *See, e.g.*, VALIAN, *supra* note 2, at 314 ("A major component of many programs aimed at increasing diversity in the workplace is awareness training, in which people are encouraged to explore together their nonconscious beliefs about gender, ethnic, and other differences."); Sara Rynes & Benson Rosen, *A Field Survey of Factors Affecting the Adoption and Perceived Success of Diversity Training*, 48 PERSONNEL PSYCHOL. 247, 247 (1995) (surveying 785 corporate diversity training programs); Jack Gordon, *Different From What? Diversity as a Performance Issue*, TRAINING, May 1995, at 25, 25-26 (describing diversity sensitivity training programs). On

reforms in the personnel system or workplace practices that are designed to increase the representation of protected groups or to reduce differences in career outcomes—measures that will be referred to here as "diversity action" programs.[162]

Diversity awareness programs might bring about a disjunction between expected liability and realized social benefit because they have the potential to create the *appearance* of "doing something about the problem" without actually reducing the targeted harm or without reducing it in proportion to the resources expended. If fact-finders are sometimes or often convinced that diversity training programs really work—that is, that they really decrease the extent to which employers are influenced by unconscious categories in judging employees—then implementing such programs might decrease the chance that a company or its supervisors will be judged in litigation to have discriminated against workers unconsciously.[163] If so, a company's investment in these programs might be worthwhile from the enterprise's point of view. But these programs may function as a form of "pseudo-precaution": there is virtually no credible evidence that diversity education programs reduce the incidence of unconsciously biased decisionmaking.[164] And the programs are not

---

workplace diversity initiatives generally, including educational awareness training, see Linda S. Gottfredson, *Dilemmas in Developing Diversity Programs*, *in* DIVERSITY IN THE WORKPLACE: HUMAN RESOURCES INITIATIVES 279 (Susan E. Jackson et al. eds., 1992).

162. These programs run the gamut of measures employers adopt with the goal of increasing the number and success of persons from underrepresented groups. Strategies include "restructur[ing] jobs and benefits [packages] in order to reduce conflicts between work and family," directing special career development efforts such as mentoring or networking programs towards female and minority candidates, creating numerical hiring targets and stepping up recruiting of employees from protected groups, and tying executives' or managers' evaluations and compensation to their "success in meeting diversity goals." Gottfredson, *supra* note 161, at 284-85. On the practice of tying managers' pay raises and bonuses to their "commitment to diversity" as a way of making supervisors more "accountable" for increasing diversity in the workplace, see MARY J. WINTERLE, WORK FORCE DIVERSITY: CORPORATE CHALLENGES, CORPORATE RESPONSES 30 (1992); Susan E. Jackson, *Preview of the Road to Be Traveled*, *in* DIVERSITY IN THE WORKPLACE, *supra* note 161, at 60; *see also* VALIAN, *supra* note 2, at 319-21 (describing a multi-pronged program at Johns Hopkins Medical School for increasing the number of successful female faculty members). For additional discussion of the Johns Hopkins program, see *infra* Part III.

163. This disjunction effectively depends on the ability to create the illusion of reducing the targeted harm—unconscious bias—without actually doing so. It therefore depends on the judicial fact-finder's inability to distinguish between "pseudo-precautions" and genuine precautions in some cases. Although companies might perceive diversity programs as helping to reduce liability exposure, the author is aware of no systematic study that demonstrates that companies with diversity programs are less often held liable for discrimination or are otherwise protected from large payouts for discrimination against workers. This issue would be a fruitful area for empirical investigation.

164. There is a dizzying amount of literature on diversity or "awareness" training, but few rigorous attempts to document its effects. In a recent comprehensive survey of the theory and practice of promoting diversity in the workplace, the author reports that "[p]opular business books on the topic have tended to be consultant testimonials and 'how-to-do-it' cookbooks." FREDERICK R. LYNCH, THE DIVERSITY MACHINE: THE DRIVE TO CHANGE THE "WHITE MALE WORKPLACE" 14 (1997); *see also id.* at 7 (There simply is "no systematic proof that diversity management programs decrease ethnic and gender tensions while increasing profits, productivity, and creativity."). Another recent review observes "that the metrics most often cited as real evidence of the success or failure of corporate 'diversity' programs are exactly the ones used to gauge the

without cost. If there is no assurance of producing the result that is the objective of the liability system, the programs could as well produce net losses as gains.

Diversity programs in the form of educational seminars and ethnic sensitivity training are frequently directed at encouraging supervisors and personnel managers to question the validity of their judgments and to suspect their own motives.[165] But, as already explained, individual employees do not know how to alter or neutralize the influence of cognitive stereotypes, and enterprises do not know how to redesign personnel systems to minimize the danger that supervisors will act on subjective bias. In light of these insights, at least one psychologist has suggested that not only are educational programs aimed at "teaching" relevant actors to make less biased judgments unlikely to succeed, but they are as likely to exacerbate the problem of unconscious stereotyping as to alleviate it.[166]

As noted, employers also have the option of adopting so-called diversity action programs aimed at altering the way employees are selected or judged or the way the workplace is organized. The potential of diversity action programs to reduce the employer's risk of being held liable for unconscious disparate treatment is a function of a unique feature of the relationship between discriminatory, actionable workplace harms, and employers' adverse actions against employees generally. As already noted, the universe of "injuries" employees might suffer as employees, in the form of decisions that detract from status or pay, is entirely internal to the workplace itself. In contrast with other types of workplace injuries, adverse actions against employees from any cause, whether discriminatory or not, are completely within the employer's control. The employer can choose, for any reason whatsoever or for no reason at all, to fire, hire, promote, demote, or otherwise offer or withdraw a desirable incident or benefit of employment. He can manipulate the background incidence of harm predictably and at will. The employer may not be able to control unconscious bias-in-the-process, but he can control the baseline incidence of adverse decisions against persons in disfavored groups to which it is compared. In contrast, an employer cannot *fully* control whether, for example, an employee contracts cancer. He can potentially control cancer risks generated from within the

---

progress of affirmative action: How many minorities and women have been hired, how many have been promoted, and to what ranks in the hierarchy?" Gordon, *supra* note 161, at 29; *see also id.* at 28 (suggesting that diversity training, by encouraging the repetition of common stereotypes about groups, does not reduce prejudice and may "reinforce prejudiced attitudes"); Rynes & Rosen, *supra* note 161, at 247, 266 (evaluating programs in terms of employees' perceptions of "training success" but failing to include any objective or standardized measure of success); *id.* at 264 (noting the paucity of "long-term evaluations" of the effects of diversity awareness, and evidence of "modest success rates for diversity management efforts in general"); VALIAN, *supra* note 2, at 314-15 ("The effectiveness of [diversity training] programs is unknown because there are no established methods of evaluation.").

165. One commentator states that the general aim of the programs is to "encourage people to be conscious of and responsive to a wide range of people who are different." ANTHONY PATRICK CARNEVALE & SUSAN CAROL STONE, THE AMERICAN MOSAIC: AN IN-DEPTH REPORT ON THE FUTURE OF DIVERSITY AT WORK 92-93 (1995); *see also id.* at 105-07; WINTERLE, *supra* note 162, at 22-23.

166. *See* VALIAN, *supra* note 2, at 315 (stating the view that "[i]n the case of gender, . . . awareness training may be counterproductive" because it makes gender schemas more salient without any evidence that this leads to a reduction in actual decisionmaking bias).

workplace, but has little control over the background level of disease to which his employees' cancer rate will be compared.

Diversity action programs can take myriad forms, running the gamut from outright "affirmative action" or "quotas," to giving weight to race or sex as a "plus factor," to a fundamental transformation of corporate organization, customs, or personnel practices.[167] Many such programs will, as a general matter, have the effect of reducing the number of unfavorable actions taken against employees in some groups relative to others. In effect, these programs will manipulate the denominator in the liability equation—that is, the number of adverse events overall. But not all such programs will necessarily affect the magnitude of the numerator—that is, the decisions that are affected in some way by unconscious bias in the process. For example, a program might boost the prospects of minority or female workers who were never targets of discrimination in the first place—thus reducing the total number of unfavorable decisions taken against members of protected groups—while bypassing individuals who *had* been true victims of unconscious bias within the enterprise. On this scenario, the employers' response will do nothing to reduce unconscious discrimination within the firm at all.

To be sure, the diversity action programs described here, including outright affirmative action, may end up preventing adverse unconscious bias from influencing *some* decisions. For example, an employer might permit some "best" minorities or women in the pool, who would otherwise be undervalued because of unconscious disparate treatment, to jump ahead in the queue in a way that advances them nearer to the position they would have occupied without bias. In the most extreme case, an employer can override any possible discrimination against minorities and women simply by refraining from taking any adverse steps against anyone belonging to these groups. Such measures do not operate by purging bias from the process of judgment itself; rather, they are better described as canceling any effects of unconscious group-based biases by disturbing the link between potentially tainted processes of judgment and outcomes.

But this strategy will necessarily entail the introduction of deliberately imposed group-conscious *compensatory* biases into the decision-making process. This type of response is problematic for two reasons. First, although the employer's preferential treatment might advance persons precisely to the position they would have occupied absent unconscious bias, it is unlikely that this will happen. As already argued, the inability of supervisors, or their employers, to calibrate attempts to compensate for unconscious biases means there is no reason to believe that the targeted harms will be precisely canceled. Second, the closer an employer's reaction gets to outright affirmative action, the stronger the argument that it constitutes a fresh violation of the antidiscrimination laws and thus a new source of potential

---

167. There will inevitably be some degree of controversy over which programs are fairly denominated "affirmative action" or whether that designation is unduly restrictive (and inflammatory). Clearly the programs vary in the degree to which they are overtly and self-consciously race or sex based, with some measures designed only indirectly to attract and retain a more diverse workforce as part of a broader set of changes aimed at increasing the productivity and satisfaction of all workers. *See, e.g.*, Gottfredson, *supra* note 161, at 284.

liability.[168] Antidiscrimination laws outlaw not just discrimination against certain disfavored groups, but rather differential treatment of any kind "because of" protected traits. Thus, an employer's attempt to avoid liability by treating some groups more favorably than others would appear to count as a "harm" within the liability system established by the literal terms of Title VII and like provisions.[169]

The important point for our purposes is that although compensatory diversity action measures could possibly issue in real "harm reduction" by overriding or reversing adverse workplace decisions against some workers whose evaluations would otherwise be negatively affected by bias, there is no necessary connection between the employer's response and the level of bias, positive or negative, that is eliminated or remains in the workplace. There is thus no systematic relationship between expenditures on precautions against liability and the costs saved through effective harm reduction, where the harm at issue is precisely race or sex-based unconscious discrimination. Given the ease of "substituting" away from the goal of decisional neutrality and the difficulty of achieving that objective, it is less likely that each decisionmaking agent will respond to a diversity initiative by engaging in more neutral processes of judgment than that each will end up implementing his or her own private affirmative action program by in effect using race or sex as a "plus factor" in an attempt to achieve a seemingly more even-handed result.[170]

If these arguments are accepted, the issue of whether inducing these types of precautions will increase net social welfare requires addressing these questions: Is the widespread adoption of diversity action programs likely to represent a cost-effective response to the threat of liability for unconscious disparate treatment? And, is such a liability system likely to be the cheapest or best method for effecting the widespread adoption of these types of programs in the workplace?

We have already seen that there is no evidence that the types of workplace diversity action programs that might be adopted under threat of liability for unconscious discrimination will produce social benefits in the form of harm reduction that accurately reflect the employer's decline in expected liability costs. A straightforward economic analysis, grounded in liability theory, predicts that the

---

168. Although some would defend workplace preferential treatment as a necessary remedial counterweight to workplace bias, scholars calling for a more vigorous attack on unconscious bias in the workplace and elsewhere do not generally choose to justify greater activism by predicting more widespread adoption of affirmative action programs. Rather, the objective is a system that is "more fair" or "more neutral" because race or sex will make less of a difference, not more. The point is that such a system is extremely difficult to achieve without introducing elements that, in themselves, deviate from neutrality.

169. This point leaves to one side the vexed question of whether and what kind of group-conscious measures can be adopted pursuant to a judicially imposed remedy for violations of antidiscrimination laws. Rather, the question is whether prophylactic trait-conscious practices on the part of employers can be regarded as increasing or decreasing the "harm" targeted by laws against discrimination in the workplace.

170. It is tempting to think that if liability spurs more searching, individualized, and careful assessments of workers, or leads to the collection of more and different kinds of information, it cannot help but have a salutary effect on the decisionmaking process. But, as already discussed, there is little support for the view that this type of effort makes for a "better" decision or for one that is consistently less influenced by trait-based biases. *See supra* notes 14-29 and *infra* notes 185-90 (discussing the effects of more information and "clinical" versus "actuarial" decisionmaking).

amount that firms expend will exceed the social savings generated by a reduction in the particular targeted harm. That is because, as argued above, there will be no necessary or proportional relationship between what employers are willing to spend on "precautions" against liability, and any reduction in the harm (unconscious bias) targeted by the liability system. Yet once that correspondence is disturbed, efficient cost avoidance cannot be guaranteed.

This point is illustrated by considering some especially popular and seemingly benign programs that employers could adopt as part of their "diversity action" programs to improve minority prospects: outreach, recruitment, training, and education directed at locating a promising pool of minority or female employees combined with extra pre-job or on-the-job training. Although costly, these measures could potentially have salutary social effects or even be socially beneficial overall.

Once again, it is difficult to justify extending Title VII-like liability to unconscious disparate treatment on the ground that it will spur the adoption of outreach and training programs. First, there is reason to believe that employers will end up disfavoring these particular strategies, since they are as likely to enhance exposure to liability as to reduce it.[171] As for trying to improve the performance of existing employees, there is little reason to think that adopting these measures in the face of a liability threat would issue in a socially optimal result. More training for women and minorities will not necessarily issue in a direct reduction in the very harms (of unconscious bias) targeted by the liability system that is driving those measures. Once again, a disparity could easily develop between the amounts expended by employers on workplace reforms and special training in the hopes of reducing their own expected liability and the social benefits of those expenditures, because investments in "precautions" will be geared to estimates of expected liability that are unrelated to actual harm reduction. Under those circumstances, the liability system cannot guarantee the cost-effectiveness of measures taken in response to it. Finally, the heavy transaction costs of a system that depends on individualized litigation makes it a poor vehicle for encouraging better training for minorities. That goal could better be accomplished by creating programs that advance the objective more directly.

This discussion does not rule out the possibility that diversity programs could, either individually or in the aggregate, generate positive social benefits net of social costs, even taking into account the costs of running a liability system. If that were so, however, that fact would be a lucky accident. It would not be a feature that necessarily followed from the design of the liability system itself. On the other hand,

---

171. Because hiring is not a rich source of potential liability for workplace discrimination, and because it is the one stage of the employment process in which employers are most likely to find it cost-effective to rely to some extent on objective criteria, employers are unlikely to elect to respond to liability by expending great efforts in recruiting more minority employees. Indeed, such increases are to be avoided because they would potentially expose the employer to enhanced liability for adverse decisions against employees once hired. *See* John J. Donohue III & Peter Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983, 984 (1991); *see also* Ian Ayres & Peter Siegelman, *The Q-Word as Red Herring: Why Disparate Impact Liability Does Not Induce Hiring Quotas*, 74 TEX. L. REV. 1487, 1489 (1996) (explaining how the threat of disparate treatment and disparate impact liability "blunts the positive incentives to hire minorities that Title VII was originally supposed to create").

the aleatory quality of that result suggests that the opposite could as well be true: overall social costs could outweigh gains. Determining which of these situations in fact obtains—and whether the proposed liability system would ultimately be worthwhile—requires nothing less than a comprehensive evaluation of the costs and benefits of the full range of workplace diversity programs generally, which cannot be undertaken here. The balance of those effects, however, will quite likely depart significantly from the situation that would prevail if the goals of the liability system—ideal race and sex neutrality—were realized.[172] Even apart from process costs, diversity programs potentially carry a number of social costs that would not be incurred in a truly color-blind system, such as decreased productivity from possible worker-job mismatch, losses to workers displaced by others, detrimental incentive effects, and social resentment and strife.[173] To be sure, even overt group-conscious measures could conceivably have some social benefits, which cannot be ignored in assessing the cost-benefit balance.[174] But the point is that if these

---

172. Depending on design and implementation, different workplace diversity and affirmative action plans will produce dramatically different outcomes for individuals within a targeted group. For example, an employer's decision to respond to a racially lopsided workforce by taking the best candidates from racially segregated lists will produce a different mix of outcomes than eliminating neutral criteria with disparate racial impact. *See* Shelly J. Lundberg, *The Enforcement of Equal Opportunity Laws Under Imperfect Information: Affirmative Action and Alternatives*, 106 Q.J. ECON. 309, 323 (1991). And neither of these scenarios will necessarily match the pattern produced by using race-blind criteria.

173. For a discussion of the perverse incentive effects of affirmative action, see Stephen Coate & Glenn Loury, *Antidiscrimination Enforcement and the Problem of Patronization*, AM. ECON. REV., May 1993, at 92, 95-97 (arguing that affirmative action tempts beneficiaries to reduce their level of effort and preparation). On some of the potential costs of other forms of diversity action programs, see generally Gottfredson, *supra* note 161, at 288-89.

Would diversity programs that include some race- or sex-conscious affirmative measures produce a more productive workforce than a move towards universal neutrality in selection criteria? The answer is complicated by the possibility that a move from the current baseline to neutrality might *not* lead to better worker-job matches and more accurate productivity-related decisionmaking overall. Neutrality might be "worse" than an unconsciously biased baseline if most unconscious bias in the workplace is rational—that is, if it reflects accurate (and for the employer efficient) generalizations about group attributes. Alternatively, there is at least some basis for arguing that neutrality would still represent a positive move overall: negative externalities from statistical discrimination might make its elimination socially beneficial. *See supra* text accompanying notes 26-33 and *infra* text accompanying notes 251-58 (discussing rational versus irrational bias and negative externalities from statistical bias). Regardless of whether a particular form of inadvertent bias is efficient or inefficient, the point is that attempts to correct it could as well prove more costly than less, or make things worse rather than better.

174. *See, e.g.*, Selmi, *supra* note 1, at 1296-1308 (arguing that affirmative action programs enhance productivity); Cynthia L. Estlund, *The Workplace in a Racially Diverse Society: Preliminary Thoughts on the Role of Labor and Employment Law*, 1 U. PA. J. LAB. & EMPLOYMENT L. 49, 59-60 (1998) (discussing the "contact hypothesis" which suggests that interactions among members of diverse groups in the workplace or elsewhere will facilitate reductions in prejudicial attitudes and might enhance cooperation); Lisa E. Cohen et al., *And Then There Were More? The Effect of Organizational Sex Composition on the Hiring and Promotion of Managers*, 63 AM. SOC. REV. 711, 719-24 (1998) (reporting data suggesting that women are more likely to be hired and promoted when an ample number of women hold jobs above and below the relevant promotional levels). *See generally* NORMAN MILLER & MARILYNN B. BREWER,

programs are implemented in response to expected liability that is not calibrated to reflect the balance of social benefits from the tortfeasor's response, there is no assurance whatever that the liability system will produce a socially optimal or even net positive result, and some reason to believe it will not.

When the process costs of running the proposed liability system are added to the mix, the picture potentially changes for the worse: Even if the widespread adoption of diversity action programs in the workplace would prove socially beneficial in the end, the key question is whether a costly, time-consuming, and resource-intensive liability scheme is the best way to spur the adoption of such programs. An individualized liability system directed at an intransigent and elusive social phenomenon is unlikely to be the most desirable way to obtain this outcome, because the same result can probably be achieved, with far less elaborate or cumbersome machinery, by direct regulation or by other more straightforward means.[175]

In sum, because unconscious bias is particularly resistant to any workable precautionary strategy or technological fix, employers will consistently find it both cheaper and easier to substitute away to measures other than "genuine" precautions—that is, arrangements not reliably known to produce a more neutral and fairer workplace. But some of these compensatory steps—such as diversity education programs—might carry significant social costs without any evidence of proportionate benefits in the form of reductions in harm. Others—such as diversity action or affirmative action programs—might reduce the effects of actionable bias for some workers, but will do so erratically and at a social cost that could well outweigh the investment in program costs and processing. It may be that certain programs designed to increase diversity in the workplace do in fact result in the abatement of the kinds of cognitive biases that give rise to distorted decisionmaking, and that programs that contain some race or sex-conscious element may ultimately have the effect of making those categories less important to decisionmaking. There is simply no way of knowing, at this point, whether that will occur.

c. Objective Assessments

The type of unconscious disparate treatment described in this Article is a problem that infects appraisal systems that rely on subjective, discretionary judgments. But employers have the option of reducing reliance on those judgments by substituting more objective methods. There are several potential pitfalls to this strategy that limit its usefulness and make it potentially costly for society as a whole. First, eliminating all subjectivity from personnel assessment would not only prove quite difficult, but would deprive employers of valuable information and vital flexibility that is considered essential to the effective management of every workplace. This point applies especially to evaluating workers' performance on the job. If claims of

---

GROUPS IN CONTACT: THE PSYCHOLOGY OF DESEGREGATION (1984) (discussing the potential prejudice-reducing effects of between-group interactions).

175. *See infra* Part II.D (stating a parallel argument on how the assignment of compensation for unconscious bias can be expected to resemble patterns seen in affirmative action programs). For additional discussion of the desirability of imposing liability for unconscious bias as a spur to the adoption of workplace diversity programs, see *infra* Part III.

unconscious bias continue to reflect current patterns of litigation, most will involve challenges to on-the-job assessments of existing employees or ex-employees rather than to the selection of new hires. But devices like tests or quantitative measures of employee performance, although not without their uses, are generally regarded as inadequate to capture the full range of information considered relevant to evaluating active workers.[176]

Second, employers will be reluctant to respond to the threat of liability by shifting to objective methods of evaluating workers because those practices carry their own threat of liability. Such methods will often have a disparate impact by race or sex, which can give rise to lawsuits under the disparate impact theory in Title VII. To escape liability, the employer must take on the notoriously difficult, uncertain, and expensive task of "validating" the selection method by proving that it is related to productivity.[177] This will discourage employers from switching to objective methods, even if they also risk liability by failing to take that option.

d. Evolution

Perhaps the most important objection to the conclusion that strict liability will not efficiently deter unconscious bias proceeds from an alternative account of how unconscious bias might abate under a threat of liability. The argument that neither well-meaning supervisors, nor well-intentioned employers, nor cognitive science, nor industrial psychology, has supplied, nor is likely to supply, a reliable program for the elimination of subjective group-based bias assumes that this type of program is necessary to that end. But that assumption ignores the possibility that change might come about through a process that resembles organic evolution. First, liability could operate as a selective tax on firms that engage in unconscious discrimination by making them pay more than others that do not maintain a flawed personnel system. That effect would operate even though no one knew how to design an unbiased system through targeted manipulation.[178] Rather, it would take advantage of chance variations in the degree to which different firms engaged in the offending behavior. The most biased firms would be placed at a competitive disadvantage by being charged with the costs of their discriminatory harms.

An alternative scenario posits that firms are more or less uniform in their degree of unconscious bias to begin with, but that practices within firms will differentiate as firms experiment with a range of new personnel methods in response to the threat of liability. Once those variations arise, the winnowing effect of liability would come into play to ensure that the firms that had by chance adopted the least bias-infected

176. *See* discussion *supra* Part I.A & note 96; *see also* FAIRNESS IN EMPLOYMENT TESTING, *supra* note 130 (discussing employment testing); Kelman, *supra* note 30 (same).

177. *See* Kelman, *supra* note 30, at 1159. *See generally* FAIRNESS IN EMPLOYMENT TESTING, *supra* note 130.

178. *See* Clayton P. Gillette, *Lock-In Effects in Law and Norms*, 78 B.U. L. REV. 813, 840 (1998) ("Evolutionary theory does not require conscious calculation by individuals about which path is superior. Rather, it suggests simply that those who set out on a path that turns out to be superior will be more successful than those who select, for whatever reason, a different path."); *see also* RICHARD DAWKINS, THE BLIND WATCHMAKER *passim* (1986); DANIEL C. DENNETT, DARWIN'S DANGEROUS IDEA *passim* (1995).

methods survived while others were driven out of business. Once again, just as organisms can evolve without possessing knowledge of how to remake themselves in ways better "adapted" to their environment, this account suggests how personnel systems can move towards neutrality under the threat of liability without anyone possessing knowledge of how to purge bias from the system.

Although arresting, this story is fundamentally flawed because it depends on a number of unproven or questionable assumptions. First, it assumes clear and sharp differences in the degree to which firms' personnel practices are infected with unconscious racial or sexual bias. But such pronounced variations among firms will not necessarily arise. Although there is evidence of individual variation in susceptibility to unconscious stereotyping, the sources of those differences are mysterious.[179] There is no reason to believe that some firms will regularly attract more bias-prone supervisors than others. Moreover, the degree of variation among individuals in the tendency to categorize unconsciously may not be large enough to drive selection among firms. Because bias in evaluations may reflect habits of thinking that are quite deeply rooted and widespread in the population at the current stage of social development, there may be rather minimal variations in inadvertent bias even across different workplace organizations and personnel systems.

Second, even if employers can be expected to change their ways in response to liability, the evolutionary approach potentially founders on the assumption that employers will move towards greater neutrality in employee evaluation. But, as already discussed, the peculiar structure of liability for discrimination may enable employers to reduce liability exposure by means other than adopting unbiased personnel methods. If diversity action programs or other responses promise a cheaper and more profound reduction in exposure to liability than "genuine" precautions against bias, the employer will take the path of substituting away from real risk reduction.

Third, differences between firms may be small and undetectable because, as already discussed, the "signal" to "noise" ratio may be very low.[180] Instead of everyone being biased, and to the same degree, perhaps only a few supervisors or decisionmakers are biased, or their biases operate only some of the time, or they

---

179. *See* Bargh, *supra* note 11, at 376; *see also supra* note 133 (discussing the influence of environment on stereotyping).

180. This observation provides one possible answer to the question of why evolution towards neutrality does not occur spontaneously without the need for the selective pressure of legal liability. Even if unconscious bias is costly to the firm, *see* discussion *infra* Part II.C.6, the effects of its operation on a firm's competitive position relative to others may be too inconsequential to make a practical difference. Alternatively, as noted in the discussion above, perhaps no firm can gain a significant edge by varying its procedures because the extent to which unconscious bias infects the range of feasible personnel practices across firms is too similar. But another explanation is that spontaneous evolution will work only if unconscious bias is irrational. *See supra* note 178. By definition, bias based on rational group generalizations will not drive firms out of business or render them less cost effective, so spontaneous competitive forces will not insure evolution away from statistical discrimination. The ratio of irrational to rational unconscious bias is unknown, although there is some evidence that a good deal of unconscious stereotyping is not wholly inaccurate. *See* McCauley et al., *Stereotype Accuracy*, *supra* note 23, at 297-99; Jussim, *supra* note 24, at 60.

influence most decisions little or not at all.[181] The operation of the evolutionary model depends on unconscious bias "making a difference" to a significant number of outcomes, which can then be used as a basis for selecting among employers. Although the frequency with which unconscious stereotyping is determinative of decisions adverse to members of protected groups is simply not known, it may be fairly low. This would provide a weak basis for any selection effect.

The final and most important objection to the evolutionary story is that a liability system targeted at unconscious disparate treatment cannot operate without a considerable amount of error. As discussed, the evidentiary and fact-finding limitations inherent in detecting unconscious bias mean that the system will almost certainly fall short in its ability accurately to identify and "tax" firms that have committed the offending conduct. Moreover, as Linda Krieger observes, judicial decisionmakers are vulnerable to the very same group or expectancy based biases in considering the evidence of workplace discrimination as the workplace decisionmakers themselves.[182] In this sense, the model of organic evolution breaks down. Evolution in biological systems is self-executing. No agent need identify the "best adapted" organisms, because environmental constraints ensure that the fittest automatically survive. But there is no guarantee of an analogous spontaneous mechanism for identifying and putting pressure on the least biased firms. Those firms are not necessarily the "best adapted" for survival.[183] The law must act to place a financial burden on the firms that engage in the undesirable practices. This requires that the legal system reliably distinguish between neutral firms and those that discriminate unconsciously. Driving discriminatory firms out of business requires the liability system to make more precise distinctions than can realistically be expected.[184]

e. Spurring Technological Innovation

The foregoing analysis exposes the weakness of a key potential rationale for holding employers strictly liable for unconscious disparate treatment: to create an incentive for employers to investigate new approaches to reducing unconscious bias in the workplace.[185] For several reasons it is unrealistic to expect employers and industrial psychologists to learn enough, even over the very long term, to develop reasonably reliable and effective strategies to combat inadvertent workplace bias. First, despite steady advances, knowledge of human psychology in general, and of the operation of human judgment in particular, is in an extremely primitive state. As already stated, an understanding of factors that are capable of moving human judgment towards greater or lesser neutrality—or greater or lesser accuracy—can

181. *See* discussion *supra* Parts I.C and II.C.1.

182. *See* discussion *supra* note 122.

183. *See supra* note 180.

184. *See supra* Part II.C.2 and *infra* Part II.D.2.b (discussing evidentiary limitations).

185. If significant risk reduction is likely to require technical innovation, strict liability is a better rule than negligence because it is difficult for courts to take speculative costs of nonobvious scientific research into account in applying a negligence standard. *See supra* Part II.B (comparing strict liability to negligence).

be gained only through sophisticated, controlled, scientific studies.[186] Constructing such studies, and gaining usable information from them, requires identifying the benchmark for accurate or neutral judgment—that is, resolving the question of "how individuals can come to make unbiased judgments" and "how judgments and behavior towards others should occur in the ideal world."[187] Alternatively, it requires comparisons under perfectly controlled conditions—that is, where the persons or situations to be evaluated differ *only* in respect of group membership or some other variable of interest. But the principal difficulty faced by psychologists studying human cognition is that the variables that influence human judgment are so complex, numerous, hard to control, and poorly understood, that a degree of scientific knowledge sufficient to yield an effective program of control and manipulation is not even on the horizon. In contrast, the technical challenges posed by other common types of workplace hazards or dangerous activities appear simple. Thus, the prospect of devising an effective technical "fix" for the problem of unconscious workplace discrimination is considerably more remote than for many other hazards.[188] Once again, there is nothing to stop employers and firms from

---

186. *See* discussion *supra* Part II.C.1-2; Wilson & Brekke, *supra* note 11, at 121-22; Grove & Meehl, *supra* note 116, at 316 (comparing "accuracy" of subjective judgments based on the availability of untainted criteria or benchmarks); *see also* Jussim, *supra* note 24, at 68 ("Identifying accuracy and inaccuracy hinges on obtaining some sort of scientific evidence") (citation omitted); Diane Kobrynowiz & Monica Biernat, *Considering Correctness, Contrast, and Categorization in Stereotyping Phenomena*, *in* 11 ADVANCES IN SOCIAL COGNITION, *supra* note 15, at 109, 111 ("[D]etermining the accuracy of a stereotype is . . . complicated. An appropriate criterion must be selected and a valid measurement tool must be created. This task may not be impossible, but it is certainly challenging."); Kruglanski, *supra* note 105, at 396 (noting that "criteria for accurate judgments are not invariably self-evident" and "[o]ften, they need to be justified by complex argument or indirect evidence," and defining three possible measures of accuracy, including (1) correspondence between a judgment and independent criterion, (2) the existence of a consensus or interpersonal agreement between judges, or (3) the usefulness of a judgment for a functional purpose, keeping in mind costs and benefits of alternatives). The difficulties of creating a benchmark for an ideal process of judgment is complicated by the fact that accuracy and "neutrality" (the absence of influence by group-based generalizations) do not necessarily go together. The theories of statistical discrimination and stereotype accuracy suggest that making use of group-based generalizations may, when information is limited, sometimes enhance the accuracy or predictive power of judgments. *See* discussion *supra* Part I.D.

187. Kobrynowicz & Biernat, *supra* note 186, at 119, 122; *see also* Trope & Liberman, *supra* note 87, at 265 (noting the difficulty of controlling for individual variations in judgment and observing that "different people may reach different conclusions depending on their motivations, on how important it is to them to avoid errors of commission and omission").

188. Risk reduction technologies for other workplace harms are usually developed in areas—like chemistry or epidemiology—in which sophisticated understandings are already in place. Techniques tailored to specific workplace problems often require fairly routine innovations or extensions of what is already known. Although the limits of science and ethical practice do generate some degree of uncertainty in these areas as well, *see, e.g.*, Wendy E. Wagner, *Choosing Ignorance in the Manufacture of Toxic Products*, 82 CORNELL L. REV. 773, 778 (1997) (noting uncertainties inherent in attempts to assess the influence of occupational hazards and other toxic sources due to "various ethical, informational, and technological constraints"), strict liability may nevertheless issue in the development of precautionary measures that are both effective and affordable, *see* Glen O. Robinson & Kenneth S. Abraham, *Collective Justice in Tort Law*, 78 VA. L. REV. 1481, 1486 (1992); Robinson, *supra* note 129, at 794.

innovating and experimenting in response to the threat of liability. But without proper scientific measures of the success of their innovations, they will be operating "blind." It will be impossible to know whether they have succeeded or not.

Finally, there are independent reasons to believe that reforms implemented primarily at the level of the individual workplace will not prove very effective. Although attitudes towards disfavored social groups have been evolving over time, these changes appear to track broad social trends and shifts in public and private discourse, images, and experience.[189] This suggests that stereotypical patterns of thought will be eroded, if at all, not through measures effected at the level of the individual workplace, but rather through a gradual sea change on multiple cultural fronts. Because broad social trends are largely out of employers' control, the use of an expensive liability system to create incentives for employers to innovate may represent a misplaced effort to induce change at the wrong level.

6. Discrimination as Unavoidable Accident

The discussion so far suggests that personnel processes in the workplace cannot be expected to progress towards greater neutrality under the pressure of a liability system targeted at unconscious group-based biases. In other words, the discriminatory "accidents" represented by contaminated assessments of employees in the workplace may be unavoidable, in the sense that there are no known effective precautions that can be taken against them.

This characterization suggests that proposed explanations for the supposed persistence of discrimination[190] in the workplace have missed an important aspect of the problem. To the extent that discrimination is "irrational," it leads to the disregard of factors pertinent to productivity in favor of arbitrary traits like race and sex. This can be expected to produce undesirable errors, or mismatches, in the selection and management of personnel. If so, firms would appear to have an incentive to eliminate discrimination, because discrimination adds to the cost of doing business and renders the firm less competitive.[191] Yet it appears that

---

189. For evidence of long-term shifts in prejudicial attitudes, see Hamilton & Sherman, *supra* note 15, at 49. The authors discuss the so-called "Princeton trilogy" in which a stereotype assessment test was administered to the general populace three times from 1933 to 1969. *See id.* Although test subjects' endorsement of some generalizations for some groups remained fairly constant over time, opinions regarding other groups—such as women—changed significantly. *See id.*; *see also* Krieger II, *supra* note 1, at 1246 (noting that employers' ability to change stereotypical reactions of supervisors and employees is limited because employers cannot control "the content of media and other cultural depictions of women and minority group members").

190. Whether there continues to be significant discrimination in the labor market is a matter of some controversy. *See supra* pp. 1131-38. *See generally* Symposium, *Discrimination in Product Credit and Labor Markets*, J. ECON. PERSP., Spring 1998, at 23 (discussing economic evidence for market discrimination against minorities and women and attributing most observed disparities to "premarket" factors).

191. Orthodox economic theory would predict that markets will be "hard on discrimination." Sunstein, *Why Markets*, *supra* note 31, at 22 ("An employer who finds himself refusing to hire qualified blacks and women will, in the long run, lose out to those who are willing to draw from a broader labor pool . . . . [B]igots are weak competitors.").

discrimination persists: employers continue in their biased ways. Markets have not eliminated discrimination.

Although commentators have tried to explain this supposed persistence in various ways,[192] one unifying theme emerges: If discrimination remains a feature of the economy, it follows that discriminators are deriving some positive benefit or utility from the practice of discrimination. These positive benefits explain why the discriminator will not necessarily choose to cease discriminating even if the "process costs" of doing so are negligible. A number of benefits from discrimination have been proposed.[193] Statistical discrimination is one example of a practice that supposedly profits the discriminator—although its effects on society as a whole are

---

192. Explanations generally fall into two types: discrimination represents the efficient working of the market. Or it is an inefficient result of market failure from negative externalities generated by practices that benefit employers at workers' or society's expense. *See, e.g.*, David Charny & G. Mitu Gulati, *Efficiency Wages, Tournaments, and Discrimination: A Theory of Employment Discrimination Law for "High-Level" Jobs*, 33 HARV. C.R.-C.L. L. REV. 57 *passim* (1998) (describing an "efficiency wage" model that is tilted against minority employees and discourages optimal work and training patterns); Donohue, *Advocacy Versus Analysis*, *supra* note 152, at 1587-91 (summarizing the theory that discriminators, like polluters, do not always internalize all costs of their activities); Paul Milgrom & Sharon Oster, *Job Discrimination, Market Forces, and the Invisibility Hypothesis*, 102 Q. J. ECON. 453, 454-56 (1987) (summarizing theories of information deficits that explain why discrimination against minorities and women persists); Selmi, *supra* note 1, at 1277-96 (reviewing theories of discrimination); Strauss, *supra* note 29, at 1621-23 (reviewing various economic explanations for the persistence of discrimination); David B. Wilkins & G. Mitu Gulati, *Why Are There So Few Black Lawyers in Corporate Law Firms? An Institutional Analysis*, 84 CAL. L. REV. 493, 514-42 (1996) (explaining discrimination within the legal profession as the product of structuring the workplace to discourage shirking and to minimize monitoring costs).

193. Richard Epstein suggests that discrimination generates positive utility by resulting in better employee-job matching, facilitating the selection and management of a more productive workforce, or satisfying some customer, worker, or owner preferences. EPSTEIN, *supra* note 152, at 28-78. Gary Becker, among others, posits a "taste for discrimination," whereby employers, customers, or co-workers enjoy positive psychic utility (or avoid negative utility) by minimizing contact with persons from minority groups. GARY S. BECKER, THE ECONOMICS OF DISCRIMINATION *passim* (2d ed. 1971); *see also* Posner, *The Efficiency and the Efficacy of Title VII*, *supra* note 152, at 513 ("For Becker, discrimination by whites against blacks is the result of an aversion that whites have to associating with blacks. This aversion makes it more costly for whites to transact with blacks than with other whites."). Becker predicts that, in a perfectly competitive market, employers with a "taste for discrimination" will be driven out of business, but firms that discriminate to satisfy customers' or coworkers' preferences will survive either as segregated firms or as integrated firms with wage discrimination. *See* Strauss, *supra* note 29, at 1633-39 (describing Becker's theory). Finally, Richard McAdams, criticizing the empirical failures of Gary Becker's "associational preference model," suggests that discrimination is best explained as a social practice that produces valuable forms of group status. Richard H. McAdams, *Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination*, 108 HARV. L. REV. 1005, 1033-62 (1995). Although maintaining a status-generating system of discriminatory practices requires effort, discrimination persists because the benefits of status production outweigh the costs of maintaining the system. *See id.* at 1063.

a matter of dispute.[194] Indeed, the theory of rational discrimination may help explain the persistence of some forms of unconscious group-based bias.[195] But that need not be the whole explanation.

The acceptance of a significant role for unconscious bias permits abandoning the assumption that persistent discrimination must bring discrete benefits to the perpetrators, either as individuals seeking psychic satisfaction,[196] or as players in an enterprise committed to matching employees to jobs,[197] or as members of a group trying to elevate or maintain group status.[198] If disparate treatment is best understood as an inadvertent byproduct of long-ingrained habits of human thought, then discrimination need bring no immediate psychic or economic benefits at all and those benefits are not needed to explain its persistence. Rather, the main obstacle to a discrimination-free workplace may simply be the difficulty of altering the mind's tendency to employ stereotypes and mental categories in social judgment.

A useful analogy can be drawn to an enterprise that suffers costly accidents (for example, injuries from plant machinery). The enterprise would ideally like to eliminate *all* such accidents. Those events bring no benefits in and of themselves, and their occurrence is pure cost or "downside" for the enterprise and its workers. But the firm may lack the know-how to design a safer workplace and developing that expertise may be out of reach. The point is that the accidents bring no benefits to the firm. The enterprise would gladly eliminate them if it could do so in a way that was consistent with its core business activities.

---

194. Even those who appear to believe that discriminatory practices are driven by positive benefits for the discriminator suggest that discrimination can generate negative externalities that decrease social welfare overall. *See supra* note 193; *see also* Donohue, *Advocacy Versus Analysis*, *supra* note 152, at 1588 (stressing that not all costs of discrimination are internalized and that "discrimination in labor markets imposes external costs that are quite analogous to the costs of pollution"); Donohue, *Is Title VII Efficient?*, *supra* note 152, at 1431 (arguing that externalities from workplace discrimination outweigh benefits); Strauss, *supra* note 29, at 1640 (suggesting that discrimination is inefficient overall because it generates negative externalities by discouraging the development of human capital). *But cf.* EPSTEIN, *supra* note 152 (arguing that benefits outweigh costs); Stephen Coate & Glenn Loury, *Antidiscrimination Enforcement and the Problem of Patronization*, AM. ECON. REV., May 1993, at 92, 92 (acknowledging the thesis that "one consequence of employment discrimination is that it harms the incentives for workers to acquire skills" because "their anticipated returns from investing in job-relevant skills are reduced," but asserting that setting lower standards through affirmative action may backfire by persuading workers they can "get desired jobs without making costly investments in skills"). The catalogue of potential social costs of "rational" discrimination includes direct losses to victims; the "discouragement" of victims leading to the underdevelopment of human capital, poor effort, or shirking; and transgenerational effects from unemployment and underemployment, including child poverty and deprivation, social isolation, anti-social behaviors, demoralization, and alienation from the culture at large. For more discussion of the assumptions underlying the identification of negative externalities of discrimination and the "discouragement effect," see *infra* Part II.C.7.

195. *See, e.g.*, Ottati & Lee, *supra* note 104, at 29, 41 (exploring evidence that some stereotyping in social interaction is grounded in accurate generalizations and thus is functional or "rational"); *infra* Part II.D.

196. *See* BECKER, *supra* note 193.

197. *See* EPSTEIN, *supra* note 152.

198. *See* McAdams, *supra* note 193, at 1086.

7. Discrimination as Avoidable Accident: The Role of Victims

The discussion so far has assumed that potential victims have no role to play in creating the risk of inadvertent discriminatory harms. It also implicitly takes for granted that the cheapest cost avoider for those harms must be the employer. This Section takes on both these assumptions. Victims of discrimination may well be in the position to exert some control over the risk that they will become the targets of unconscious bias in the workplace. Indeed, employees might be able to control that risk more efficiently than employers.

The possibility that employers and workers share control over the risk of inadvertent discrimination has important implications for the operation of a strict liability system. In general, strict liability without contributory negligence will not provide the optimal rule where the victim can affect the risk of an actionable harm.[199] If victims are insured through compensation for bias-induced harms, they will have less incentive to take effective precautions against those harms, even if they have the means to do so.[200]

How do these observations affect the issue of who should bear the costs of unconscious disparate treatment? The answer is complicated by the need to consider both the victim's potential role in triggering an adverse event for which he would be entitled to recover—that is, his role in eliciting discrimination—and his role in increasing the risk of unfavorable workplace treatment generally. With respect to the latter, the availability of "insurance" against *some* workplace setbacks (those due to discrimination) might reduce the employee's incentive to avoid behaviors that enhance the risk of unfavorable actions (like discipline and termination) generally. To be sure, that possibility depends on adopting the assumption that it is quite difficult to distinguish tortious from nontortious workplace harms—an assumption for which this Article argues.[201] Nevertheless, the magnitude and direction of the inevitable errors are very hard to predict. Unless judicial determinations of liability are chronically tilted towards false positives, or employees are overly optimistic about their prospects of winning discrimination suits, insuring employees against unconscious bias should not create significant moral hazard for misbehavior of a general kind.[202]

---

199. *See* LANDES & POSNER, *supra* note 62, at 39-40; SHAVELL, *supra* note 36, at 26-32; Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1 (1980).

200. *See, e.g.*, Richard A. Epstein, *Products Liability as an Insurance Market*, 14 J. LEGAL STUD. 645, 653 (1985) ("Individuals with insurance against certain types of losses are more likely to engage in risky conduct than those who do not have that insurance.") (citing Steven Shavell, *On Moral Hazard and Insurance*, 93 Q. J. ECON. 541, 541 (1979)).

201. *See supra* Part II.C.2.

202. The behavioral effects of the availability of a cause of action for discrimination would depend partly on the employee's probability of benefitting from an erroneous judicial finding of compensable discrimination balanced against the expected loss from the court ruling against the employee. Potential victims may perceive these probabilities incorrectly, however. If claimants are overly optimistic about the potential for prevailing in a bias suit, that would enhance the moral hazard in the system. In theory, this effect will apply regardless of whether the claim is for

It is also possible that factors within an employee's control could influence the specific risk at issue: that the employee will become the target of unconscious bias. Although the suggestion that the degree to which group-based biases influence judgments about particular individuals might depend on the victims' own attributes or behavior will inevitably strike some as repugnant "victim-blaming,"[203] that prospect must nevertheless be taken seriously in the context of this analysis. All available evidence suggests that trait-based "mental contamination" of social interactions is a highly variable and nuanced phenomenon that is responsive to contextual cues. The degree to which unconscious bias figures in a decision at all may vary widely with the characteristics of the person being assessed and the type of information that person makes available about himself. This suggests that elements that vary from person to person and are at least partly within that person's control—such as background, education, appearance, grooming, demeanor, manners, speech patterns, work habits, personal conduct, and personality type—could well determine whether and to what extent an individual elicits an unconsciously biased response. Moreover, this variation and sensitivity to context is almost surely more pronounced when discrimination is subtle or inadvertent than when it is crude, deliberate, or grounded in overt bigotry.[204] The exact circumstances that give rise to these variations, and precisely how they operate, are empirical questions that cannot be definitively answered in the current state of understanding. There is some cautious support for the possibility that reliance on race or sex-based presumptions will decline as the type or quality of credentials improves.[205] But any

---

conscious or unconscious discrimination. In practice, it depends on the potential for false positives, or for self-delusion, generated by each type of claim.

203. *See, e.g.*, KELMAN & LESTER, *supra* note 30, at 218 (describing the widespread tendency of left multiculturalism to "avoid[] any hint of 'blaming the victim,'" which is described as the view that "we can best understand the victim's problems by looking at *her* traits, rather than the traits of those who evaluate her or 'treat' her in a particular way") (emphasis in original).

204. *See, e.g.*, Krieger I, *supra* note 1, at 1310-13 (discussing the assumption in current law that manifestations of overt or intentional group-based animus are fairly consistent from victim to victim, whereas the evidence suggests that unconscious bias is more variable and context-dependent). To be sure, practitioners of overt racial or sexual bias may sometimes pick and choose their victims based on aspects of the victim's characteristics or behavior. Courts recognize that a particular employer's failure to discriminate against all members of a minority group in the workplace does not rule out sporadic animus directed against particular members of the group. But our intuition is that actors motivated by purposeful racial or sex-based animus will usually make no such distinctions. *See, e.g.*, *id.* at 1310 ("If a person discriminates against members of a particular gender, racial, or ethnic group, we expect him to do so *consistently*. We assume a particular decision maker is unlikely to have a 'taste for discrimination' one day and not the next.") (emphasis in original) (footnote omitted).

205. There is speculation that outgroup employees who conform most closely to behavioral and lifestyle paradigms of the dominant group are less likely to be the targets of bias. *See, e.g.*, Barbara J. Flagg, *Fashioning a Title VII Remedy for Transparently White Subjective Decisionmaking*, 104 YALE L.J. 2009, 2009-15 (1995) (discussing the appeal of "whiteness" and "acting white"); Kevin Lang, *A Language Theory of Discrimination*, 101 Q. J. ECON. 363 *passim* (1986) (discussing reactions to unfamiliar cadence and manner of speech); Mari J. Matsuda, *Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction*, 100 YALE L.J. 1329 *passim* (1991) (analyzing the role of speech in discrimination). The quality of credentials might also influence the extent to which employers unknowingly fall back on stereotypical

such "sliding scale" effect will depend not on anything an employer can do, but on the possibility of a "moving target"—that is, on the employee's willingness and ability to vary his own profile of qualifications and behavior. If an evaluator's susceptibility to mental contamination is indeed sensitive to victim characteristics, this suggests that employees might be able to fend off unconscious bias by manipulating aspects of their presentation or behavior on the job, or by acquiring better credentials, education, or experience. To be sure, such exertions may not completely eliminate all group-based bias, and group members may still be unconsciously undervalued. But if victims can *reduce* the degree of undervaluation by changing their behavior or acquiring different attributes—and if employees are the ones who can most effectively bring about that result—it could well make sense to let the costs of unconscious bias remain where they fall.

One objection to this notion, however, is that the power to choose how to present oneself to the world is not the same as knowing how to alter one's presentation for the purpose of *minimizing bias*. Indeed, the victim would appear to be as much in the dark as the employer on this score. Although it is possible that counterstereotypic qualities tend to deflect bias, this view amounts to little more than speculation based on very limited evidence. The fact remains that employees no more understand how deliberately to manipulate employers' unconscious thought processes than employers know how to control their own biased thinking.

But this lack of know-how does not necessarily eliminate all moral hazard for victims. Although victims and employers would appear to be equally incapable of consciously planning a constructive response to the incentives created by bearing the costs of unconscious bias, that symmetry is illusory. Consider the way in which an employee might go about trying to avoid becoming the target of unconscious discrimination. The employee will have no choice but to proceed by trial and error, since he doesn't know for sure which moves will work. But the employees' efforts will not be totally random: he will almost certainly proceed by trying to figure out how to obtain a generally favorable response from his workplace superiors. The employee will use feedback as a guide, amplifying the approaches that seem to please and abandoning those that do not. This looks like the flip side of the employer's most likely response, discussed above, to a liability rule that shifts the

---

generalizations. A black man with an honors engineering degree from Harvard, for example, might avoid triggering presumptions that would attach for someone with a degree in sociology from a lesser known institution.

The studies that seek to examine whether the influence of unconscious stereotypes varies with whether victims exemplify or defy common stereotypes provide some guarded support for the view that inadvertent biases are less dominant under "counterstereotypic" conditions. *See, e.g.*, Charles Stangor et al., *An Inhibited Model of Stereotype Inhibition*, *in* 11 ADVANCES IN SOCIAL COGNITION, *supra* note 15, at 193, 202-03 (describing, for example, how Hispanic targets observed in front of a library are described using fewer stereotypic terms than Hispanic targets observed near a bullet-riddled window); *supra* Part II.C.1 (discussing counterstereotypic information). However, there is a remarkable paucity of well-controlled studies that try to test this hypothesis for workplace-based appraisals.

The situation for gender bias may be similarly complex. The question of how women can best dodge sex-based stereotyping by workplace superiors is the subject of endless private worry as well as public speculation. *See* Price Waterhouse v. Hopkins, 490 U.S. 228, 251 (1998) (noting the double bind of acting too feminine or too masculine).

costs of unconscious bias onto the firm: Employers, like protected employees, will respond by attempting to bring about more favorable outcomes for the persons covered by the liability rule. But there is an important difference: employers can afford to be cynical, but employees cannot. As already noted, the employer has complete and dictatorial control over the full range of adverse employee outcomes, whether discriminatory or not. But employees have no parallel authority. Thus, although both employers and employees, if made to bear the costs of unconscious bias, can be expected to try to minimize the number of unfavorable employment decisions that, respectively, they make or suffer, employees are much more limited in their means of accomplishing this result. Employers have the option of treating employees well whether or not they are good employees. But for employees, the most effective way to avoid becoming a target of unconscious discrimination is probably also the most effective way of ducking unfavorable treatment in general: by striving to become a model worker.

Any effort along those lines, however, is likely increase the worker's productivity; and if the effort does not prove too costly, it will generate a corresponding increase in net social welfare.[206] To be sure, it will be impossible to know for sure whether a particular employees' efforts are diminishing the influence of categorical race or sex-based bias. Nor can it be said with certainty that behavior that elicits a more favorable employer response is necessarily more productive or beneficial. Employers sometimes favor workers for bad reasons, or for reasons only loosely related to actual performance on the job. Assessments of workers, even when free from actionable bias, are subject to errors and a range of unexplained variations.[207] But the discussion in this Section rests on the assumption that this pattern will not dominate: Absent discriminatory biases or perverse incentives to sacrifice productivity for other payoffs (such as reducing the risk of liability), it is not implausible to assume in a competitive marketplace that employees who do well in the workplace—that is, employees who manage to minimize the adverse decisions taken against them—will be among the most productive.

As we have seen, forcing employers to bear the costs of unconscious bias may result in potentially costly measures, such as diversity training, diversity action programs, or affirmative action, that are designed to reduce the risk of liability but may do very little to cure the precise problem of bias targeted by the liability system. Moreover, although it is sometimes claimed that these types of workplace reforms increase worker satisfaction and productivity,[208] there is very little solid evidence of this effect.[209] In contrast, an employee who responds to the prospect of bearing the risk of loss from unconscious bias by looking for ways to minimize his own victimization will look for ways to be a better employee. This search is likely to have a productivity-enhancing effect. If so, then imposing strict liability for unconscious disparate treatment on employers could well be ill-advised to the extent

206.On the balance of costs and benefits, see *infra* pp. 1205-06.

207. *See* discussion *infra* Part II.D.2.b on irrational "noise" or arbitrariness in labor markets.

208. *See, e.g.*, Selmi, *supra* note 1, at 1255; Estlund, *supra* note 174, at 54-58.

209. *See, e.g.*, Gottfredson, *supra* note 161, at 300; VALIAN *supra* note 2, at 314-15; Rynes & Rosen, *supra* note 161, at 247.

that it shifts losses from the party whose response is more likely to increase social welfare to the party less likely to do so.[210]

One potential problem with the analysis so far is that it gives short shrift to the cost side of the equation. Whether leaving employees to bear the costs of unconscious bias is efficient—that is, net socially beneficial—is a function not just of whether the employee will be led to increase his productivity, but what it costs him to do so. Some commentators have argued that group-based discrimination may induce socially wasteful overinvestments in education or training.[211] Alternatively, employees might incur grievous personal costs from efforts to minimize bias, including the sacrifice of individuality, identity, or choice.[212] Finally, the notion that employees will in fact respond "constructively" to unregulated discrimination could be criticized as unrealistic. An oft-repeated assertion in the discrimination literature is that group-based bias will lead to discouragement: members of the targeted group will underinvest in human capital and will underperform on the job because the expected payoff from these inputs will be less than for others.[213] Finally, the decision to take the victim's role into account in selecting a liability rule depends to some extent on justice-based and equitable considerations that transcend the terms of a

---

210. One objection to this conclusion is that the magnitude of the moral hazard from shifting the costs of unconscious bias away from victims to employers may be fairly small. It can be argued that employees have substantial standing incentives to please their employers, which will swamp any incremental effects of making them bear the costs of unconscious bias. There is no easy way to answer this objection, because not enough is known about the relative influence of legal protections and extra-legal factors on employee effort and performance. The possibility that the liability rule for bias will have some significant effect, however, cannot be dismissed out of hand. An employee's belief that subconscious bias is potentially an important factor in the workplace and can be manipulated could significantly influence that employee's pre-market and workplace choices.

211. *See* A. MICHAEL SPENCE, MARKET SIGNALING: INFORMATION TRANSFER IN HIRING AND RELATED SCREENING PROCESSES *passim* (1974); A. Michael Spence, *Job Market Signaling*, 87 Q. J. ECON. 355, 368-74 (1973). See also Charny & Gulati, *supra* note 192, at 78-83, for the argument that victims caught up in discriminatory systems will invest in educational signals that improve the victims' prospects in the workplace but fail to increase productivity in a cost-effective manner.

212. The strategies could range from simple adjustments in attitude, dress, and demeanor, to more problematic shifts in speech patterns or lifestyle choices that are central to cultural identity, solidarity, or individual personality.

213. See *supra* note 194 for a discussion of the "discouragement" effect. This idea often surfaces in discussions of statistical discrimination, and is used to defend the conclusion that although rational bias may be efficient for the employer it induces negative externalities that are inefficient for society as a whole. See *supra* notes 192 and 194 for a discussion of externalities. For examples of this view, see Arrow, *supra* note 29, at 96 ("If the employer is going to judge by race, then there is no reward for [greater] investments. They will not be acquired, and then the statistical judgments will be confirmed."); Darity & Mason, *supra* note 127, at 84; Williams & Sander, *supra* note 130, at 2035 (suggesting that race-based statistical discrimination is self-reinforcing because the discouragement and resentment that results from being judged by group norms causes targets to underperform). *See also* Strauss, *supra* note 29, at 1639-43; Sunstein, *Why Markets*, *supra* note 31, at 29-31. For an economic model of statistical discrimination that describes reduced payoffs for workers, see Edmund S. Phelps, *The Statistical Theory of Racism and Sexism*, 62 AM. ECON. REV. 659 (1972).

cost-benefit analysis so narrowly conceived. We may be uneasy about placing the onus for finding strategies to reduce unconscious bias on potential victims when they do not "deserve" to be the targets of risk in the first place. We may ultimately judge the employer preferences that employees must satisfy in their quest to minimize bias as unreasonable, culturally arbitrary, and only tenuously related to any legitimate "business purpose."

Considerations of justice and employer hegemony, although important, are beyond the scope of this Article. Leaving aside these difficult issues, however, one answer to the concern that victims will make overly costly sacrifices or investments is that victims generally will not pursue particular options designed to reduce their own risks of victimization unless it is worth their while: those measures must promise to generate *net* benefits to the victims themselves. Since victims will bear most of the costs of trying to forestall their own victimization and will also reap the lion's share of rewards from success, they are probably a good judge of whether a particular response is socially beneficial overall. In other words, the employee's internalization of most significant costs and benefits of any change in strategy should operate as a hedge against wasteful overinvestment in most cases. Although expecting employees to bear the costs of making changes in response to the prospect of being victimized by unconscious bias may not be fair, it is unlikely to be inefficient.

As for the contention that victims may *not* internalize all costs of investments in response to discrimination (most notably some "signaling" costs),[214] that view depends on the assumption that investments in signaling systems may not always be socially efficient. The signals used to sort workers (such as education) do not necessarily enhance productivity. Rather, they may simply reveal pre-existing differences in talent or ability. But there is no a priori reason, at least from an economic point of view, to distinguish the costs of matching and sorting employees from other factors bearing on their contribution to the net profitability of an enterprise.[215] Although signaling practices may occasionally generate wasteful equilibria,[216] the employer's preference for employees who possess relatively reliable signals of quality or ability will more often reflect real net benefit to the enterprise and to society as a whole.

Finally, the prediction that discrimination that effectively discounts the inputs of members of some groups will invariably lead to "discouragement," reduced effort, and less investment in human capital is almost surely wrong. More likely the true picture is far more mixed, with different people responding in different ways. Discrimination can be viewed as a form of adversity, in the sense that it can reduce payoffs to individuals for a fixed amount of effort. But victims of adversity are always faced with a range of options. Specifically, they can decide whether to exert greater or more costly efforts in the face of reduced payoffs or to decrease effort. Whether the selection of one or the other strategy is "rational," either from the victim's point of view or in some broader sense, does not admit of a categorical

---

214. *See, e.g.*, SPENCE, *supra* note 211; Charny & Gulati, *supra* note 192, at 57.

215. *See* KELMAN & LESTER, *supra* note 30; Kelman, *supra* note 30, at 1203-04 (distinguishing between "gross" productivity, which reflects work output and quality only, and net productivity, which represents an employee's work output minus all expenses related to hiring, managing, and supporting the worker).

216. *See, e.g.*, SPENCE, *supra* note 211.

answer. Whether it is "rational" for the victim himself depends on characteristics peculiar to each individual, including the relative dominance in each person's preference structure of what labor economists would term the substitution effect and income effect from unfavorable changes in that person's economic prospects.

The substitution effect is the product of the "price" or opportunity cost of leisure. As the payoff to work decreases (as with discrimination), leisure costs less in forgone earnings, which causes work effort to decline. The income effect, on the other hand, is the product of the decrease in the demand for a good (for example, money) as income rises, or an increase in demand as income falls. The income effect will lead a person who is poorer because of discrimination to work harder to restore the level of income he otherwise would enjoy.[217] In the case of discrimination, these effects work in opposite directions, and whether a person will respond to discrimination by working harder or "smarter," or working less hard, depends on the relative dominance of the effects for that person.

The mix of these two effects is not constant for everyone: it varies with the taste for leisure over material gain, the consumption value or moral value of work, and other factors relating to discount rate, long term and second-order goals, group identification, work-ethic, self-concept, personality, cultural values, and normative expectations. The complexity of these factors means that not everyone will deal with the prospect of discrimination by becoming discouraged, studying less, or working less hard. Some may have the opposite response. It could be argued that groups in which most persons respond to discrimination with more effort rather than less will be more successful in the long run. Indeed, if unconscious bias turns out to vary regularly with factors victims can manipulate, individuals who counter discrimination with redoubled effort or investment may well place themselves at a distinct advantage relative to those who do not. In light of these observations, it should be clear that victims of unconscious bias are not doomed to a downward spiral of employment-related losses if unconscious discrimination goes uncompensated. Rather, there may be much that potential victims can do to lighten the burden placed on them by this unfortunate phenomenon.

Even if victims are not completely helpless against unconscious bias, however, benefits will not be forthcoming if most people simply refuse to make a positive response. Advocates of leaving the costs of unconscious bias where they fall must confront the possibility that most persons who believe themselves victims will react as the dominant theory predicts: Some will become discouraged and decrease their amount and quality of effort, while others will conclude that the cost of adjusting behavior or increasing investment is simply too great and overwhelms any potential gain. If the system must accept people and their preferences as they find them, this is a potent objection. Nonetheless, not only is there no reason to believe that the balance of income and substitution effects are uniform across persons, there is also no reason to believe that they are immutably fixed. The preferences that determine how individuals deal with discrimination almost surely can be altered by reason and

---

217. *See, e.g.*, DANIEL S. HAMERMESH & ALBERT REES, THE ECONOMICS OF WORK AND PAY 34 (3d ed. 1984). For an instructive application of the concepts of income and substitution effects to the analysis of how individuals will respond to legal protections for the learning disabled, see KELMAN & LESTER, *supra* note 30, at 188-94.

persuasion as well as by shifts in norms, attitudes, expectations, and cultural values. Persons who at first accept discouragement as an "appropriate" or rational response to discrimination might come to see this as counterproductive or self-defeating. Others who initially regard certain behaviors as central to cultural identity or individuality may yet come to see compromise as ultimately less costly or more worthwhile than maintaining a personal status quo.

Even if these arguments are accepted, it may nevertheless strike many as perverse to structure a legal regime to "reward" victims who can manage a constructive response to adversity rather to tax the employers who are creating the adversity. But that stance assumes both that employers can be taxed accurately and effectively and that shifting costs through the legal system to reduce unconscious bias will actually *work*. Although trying to hold victims harmless may make occasional victims better off, the overall effect could well be to increase net social costs by discouraging the cheapest cost avoiders (that is, victims) from making the most socially constructive response to loss. Leaving the costs where they fall may represent the best we can do in an imperfect world.

D. Compensation and Insurance Against Unconscious Disparate Treatment

As argued above, employers do not know how to purge the influence of unconscious stereotypes from workplace assessments. Although employers might respond to the threat of liability for unconscious bias by implementing diversity action programs that end up reducing the number of adverse decisions against protected individuals—and which therefore may cancel the effects of some indeterminate amount of unconscious bias—the relationship between exposure to liability and reduction in the targeted harm will be irregular and unpredictable. Since imposing strict liability on employers for unconscious forms of bias cannot be expected to produce targeted deterrence, it cannot be relied upon to produce efficient cost avoidance. Indeed, it might detract from efficient risk reduction by shifting costs away from employees to employers.

But efficient harm reduction is only one possible objective of a liability system. A system might also seek to compensate victims for losses and to insure potential victims efficiently against harm. The prospect of generating considerable transaction costs without effective or efficient risk reduction must then be weighed against the liability system's potential to advance the goals of effective compensation or efficient insurance.

With respect to compensating victims, efficiency is not the only concern. Considerations of justice loom large. Among the principal goals of compensation systems has always been to return individuals to the position they would have occupied absent the wrongful conduct. Where unlawful reliance on race or sex results in concrete losses to the victim—that is, where the effect of bias in the workplace is "determinative"—there are strong arguments, grounded in principles of corrective or compensatory justice, for making victims of unconscious disparate

treatment whole for any resulting losses.[218] Although some theories of compensatory justice might attach importance to distinctions based on the state of mind of the person being held liable,[219] it will be assumed for purposes of this discussion that if losses to victims caused by unconscious bias can be made up efficiently, justice dictates that they should be.[220]

---

218. This Article is primarily about the social costs and distributional effects of imposing tortious liability for unconscious discrimination. Issues of compensatory justice are largely beyond its scope. Thus the discussion indulges a presumption that transfers from the "injurer"—here taken to be the employer—to the "victim" of discrimination—the employee—are appropriate for the purpose of making up the losses suffered by the victim. *Compare* Stephen R. Perry, *The Moral Foundations of Tort Law*, 77 IOWA L. REV. 449, 465 (1992) (noting that the Coasian view of transactions calls into question the identification of victims or perpetrators of a harmful interaction, since the participation of both is required to bring about the harm), *with* Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151, 164-89 (1973) (asserting a non-reciprocal theory of causation of harm).

219. Conventional accident analysis makes no initial categorical distinction between losses inflicted "on purpose" and those incurred through carelessness or inadvertency, although the defendant's state of mind will have some bearing on the real-world cost-benefit calculus. In contrast, whether compensation for inadvertent harms would either be mandated by, or consistent with, principles of justice is a highly debated question. It is unclear whether, and which, justice-based theories would endorse compensating victims of discrimination regardless of whether the harm is inflicted intentionally or negligently, or the harm is unforeseeable, or the perpetrator could not have "taken greater care." *See* Perry, *supra* note 64 (surveying views in the compensatory justice literature on the relationship between causation, foreseeability, state of mind, and justifications for compensation); *see also* Jules L. Coleman, *The Morality of Strict Tort Liability*, 18 WM. & MARY L. REV. 259, 275-86 (1976) (expressing concern about the moral implications of tort theories that assign the costs of accidents to parties who are not in some sense "blameworthy" or "at fault"); Mahzarin R. Banaji & Anthony G. Greenwald, *Implicit Stereotyping and Prejudice*, *in* 7 THE PSYCHOLOGY OF PREJUDICE: THE ONTARIO SYMPOSIUM, *supra* note 24, at 55, 70-71 (questioning the moral implications of holding persons responsible for automatic unconscious biases that are outside individual control and arguing for "removing responsibility from individual perpetrators of social crimes of stereotyping and prejudice"); Bargh, *supra* note 11, at 364 (suggesting that "[i]f people cannot help stereotyping, then they cannot be held personally responsible for their actions, and so cannot be sanctioned for any prejudicial actions").

220. The case for compensating victims for inadvertent bias in the workplace is enhanced if the failure to make victims whole for losses resulting from workplace discrimination can be expected to have significant secondary or spillover effects that compound the victim's injury. Once again, it is difficult to assess the importance of any such factors for unconscious bias because it is hard to get a sense of how much social judgment is affected in the real world. This Article suggests that inadvertent stereotyping is probably only of sporadic or minor importance in many settings. But, if unconscious bias is a significant phenomenon—that is, it often makes a difference and group members can only do so much to minimize bias—then it may have important collective or secondary effects. In addition to depressing target groups' income, employment prospects, rates of employment, and status in the workplace, these patterns can result in a cascade of self-reinforcing consequences, including poor health, educational underachievement, alienation, caste-like stigma, and rejection of dominant norms and cultural expectations. These can in turn fuel undesirable neighborhood, community, and transgenerational effects, which in turn can depress minorities' prospects for successful employment. *See, e.g.*, MOSS & TILLY, *supra* note 131, at 30 (discussing "additional round effects" of high rates of black male unemployment); WILLIAM JULIUS WILSON, WHEN WORK DISAPPEARS (1996) (same); Cass R. Sunstein, *The Anticaste Principal*, 92 MICH. L. REV. 2410 *passim* (1994). Not only do conventional remedies for employment

The goal of compensation, however, requires that compensatory transfers be appropriately targeted. Compensation should be paid to actual victims. To the extent that the system falls short of this ideal—that is, to the extent that the assignment of compensation is inaccurate—the goal of corrective justice for individuals is vitiated. In light of this observation, the ensuing discussion addresses the following questions: How does making employers strictly liable for unconscious disparate treatment fare in advancing the goals of efficient insurance and fair, efficient, and accurately targeted compensation for victims?

1. Insurance Against Unconscious Bias: Tangible and Intangible Losses

Compensation for tangible or monetary losses from unconscious discrimination would ordinarily be expected to provide an efficient form of insurance. Assuming some risk aversion on the part of victims, an insurance system that makes up tangible accidental losses should increase victims' total well-being overall.[221] Viewing compensation for discriminatory harms from the point of view of insurance theory also sheds light on the question of whether compensation should be awarded for (conscious or unconscious) "nondeterminative" bias—that is, bias-in-the-process

---

discrimination (including front pay, back pay, and equitable reinstatement) typically disregard many of these consequential costs, but these remedies will not necessarily head off all secondary effects. Also, the emotional or dignitary harms from discrimination might linger regardless of compensation for material losses. Indeed, those ill effects might not be wholly eliminated even by adding compensation for the intangible harms themselves. See *infra* text accompanying notes 222-24 for a discussion of compensation for nonpecuniary losses from discrimination. Although the ideal would be to eliminate the bias that warrants compensation, generous compensation could at least mitigate these effects and would be worth achieving if otherwise feasible.

221. Insurance generally is thought to be "efficient" if the victim would choose to insure himself against the harm. If victims are risk averse, they will gain by transferring money from the preaccident to the postaccident state until the marginal utility of money in both states is equalized. This occurs when monetary losses are fully compensated. Persons who suffer monetary losses from accidents (either because of lost income or enhanced expenses) have a higher marginal utility of money, and will experience a gain in utility from a transfer of money from the wealthier (preaccident) to the less wealthy (postaccident) state. Victims can thus be expected to purchase full insurance for tangible losses because they will choose to transfer money until monetary losses are fully made up. The insurance itself, by abating risk, adds an extra component of positive utility over and above the equalization of wealth. *See, e.g.*, PAUL H. RUBIN, TORT REFORM BY CONTRACT 29-40 (1993); Patricia M. Danzon, *Tort Reform and the Role of Government in Private Markets*, 13 J. LEGAL STUD. 517, 520-24 (1984); Ellen Smith Pryor, *The Tort Law Debate, Efficiency, and the Kingdom of the Ill: A Critique of the Insurance Theory of Compensation*, 79 VA. L. REV. 91, 99-104 (1993); Schwartz, *supra* note 153, at 362-67.

The insurance story for tangible harms from discrimination is not without complication, however. The types of harm inflicted may sometimes give rise to severe valuation problems (e.g., as in trying to assess the monetary loss that corresponds to the employee's forgone future opportunities). Antidiscrimination liability systems also necessarily lump together into the same "insurance pool" potential victims with very different degrees of risk aversion toward the prospect of suffering losses from workplace bias, which can potentially produce inefficiencies. *See, e.g.*, Jon D. Hanson & Kyle D. Logue, *The First-Party Insurance Externality: An Economic Justification for Enterprise Liability*, 76 CORNELL L. REV. 129, 139-41 (1990).

that produces no measurable loss in concrete job benefits or monetary compensation.[222] That question is potentially quite important to the legal treatment of unconscious discrimination, since, as already discussed, inadvertent stereotyping may not make a difference to many decisional outcomes. In the absence of a tangible loss, a system that awarded compensation under these circumstances could be regarded as creating insurance for emotional, dignitary, or nonpecuniary harms—that is, for some form of "pain and suffering." There is a large, though not unchallenged, body of work that counsels against granting compensation for these types of harms, based on the prediction that compensation is not likely to improve the well-being of victims overall. The benchmark for when insurance is efficient is whether a victim would choose to purchase insurance against a loss. Economists reason that a person would not rationally choose to purchase insurance against intangible emotional harms because, once tangible losses have been made up, money is worth less to a person after the harm has been inflicted than it is in the uninjured state.[223] Should employees be compensated for nondeterminative discrimination? The orthodox theory predicts that victims of discrimination will not insure themselves against this nonpecuniary form of harm, and therefore a liability system that builds in this insurance component by awarding compensation for "bias-in-the-process" is unlikely to be efficient. In reality, however, the question of whether compensating discrimination that issues in no monetizable harm would enhance social welfare cannot be answered definitively in theory. The relative marginal utility of money in any postaccident state is ultimately an empirical question, and the answer may vary from case to case.[224] Further confounding the issue is the observation that minority group members will not necessarily function in real life

---

222. Some scholars have suggested that victims should be compensated for being exposed to biased processes or procedures, regardless of outcomes or tangible harms. *See* Mark S. Brodin, *The Standard of Causation in Mixed Motive Title VII Actions: A Social Policy Perspective*, 82 COLUM. L. REV. 292, 316-26 (1982); Mary F. Radford, *Sex Stereotyping and the Promotion of Women to Positions of Power*, 41 HASTINGS L.J. 471, 526, 528-34 (1990); Stonefield, *supra* note 27, at 134-75; Weber, *supra* note 9, at 515-24. *See generally* Lidge, *supra* note 27 (arguing that nondeterminative discrimination should be compensated). An analogy could be drawn between this argument and the suggestion that persons are entitled to compensation for "exposure to risk" regardless of whether any injury ever materializes. *See, e.g.*, Christopher H. Schroeder, *Rights Against Risks*, 86 COLUM. L. REV. 495 *passim* (1986); *see also* David McCarthy, *Liability and Risk*, 25 PHIL. & PUB. AFF. 238, 247-53, 259-62 (1995); David McCarthy, *Rights, Explanations, and Risks*, 107 ETHICS 205 *passim* (1997); Robinson, *supra* note 129 *passim*; Kenneth W. Simons, *Corrective Justice and Liability for Risk-Creation: A Comment*, 38 UCLA L. REV. 113, 114-18 (1990). For further discussion of "compensation for risk," see *infra* text accompanying notes 235-36.

223. *See* Chamallas, *supra* note 1, at 505 n.164; *see also* RUBIN, *supra* note 221, at 29-40; Mark Geistfeld, *Should Enterprise Liability Replace the Rule of Strict Liability for Abnormally Dangerous Activities?*, 45 UCLA L. REV. 611, 630 (1998); Pryor, *supra* note 221, at 101-04.

224. *See, e.g.*, Heidi Li Feldman, *Harm and Money: Against the Insurance Theory of Tort Compensation*, 75 TEX. L. REV. 1567, 1573-77 (1997); Pryor, *supra* note 221, at 125-36; Margaret Jane Radin, *Compensation and Commensurability*, 43 DUKE L.J. 56, 69-83 (1993); *see also* Stephen P. Croley & Jon D. Hanson, *The Nonpecuniary Costs of Accidents: Pain-and-Suffering Damages in Tort Law*, 108 HARV. L. REV. 1787, 1812-45, 1857-95 (1995) (suggesting that the absence of a market for insurance against nonpecuniary harms is not dispositive of the question whether consumers would want such insurance or whether it would be efficient).

as their own insurers against workplace discrimination. Employers may spread the costs of liability for discrimination among a variety of economic actors, including but not limited to victims.[225] In that case, the observation that money or other resources may have relatively low marginal utility in the postaccident state raises a question that goes to the wisdom of *any* redistributional move: "relative to what?" The issue boils down to whether compensation will result in transfers from persons with lower marginal utility to persons with higher marginal utility for the transferred resources. It is impossible to answer that question in the abstract.

Insurance theory thus offers little help in resolving the debate between advocates of compensation for discrimination that falls short of producing tangible harm and discrimination that causes concrete losses.[226] Even if compensation is restricted to tangible losses, however, serious questions of fairness are raised by viewing liability for group-based discrimination in general, and unconscious bias in particular, as an insurance system. If, as suggested above, *all* employees end up bearing the costs of the compensation system, it is unclear whether this allocation can be justified. On the one hand, such a system will transfer resources from individuals who are exposed to minimal risk—non-minority or male workers—to members of protected groups who are at far greater risk. But individuals who are known ahead of time to bear no risk are not properly included within a "risk pool" charged with self-insurance. That is subsidy, not insurance, and as such enhances the possibility of an inefficient transfer.[227] (This effect would be softened, however, if employers shift a disproportionate amount of the costs of compensation to employees who are most likely to collect compensation.) On the other hand, it can also be argued that no

---

225. The determination of whether insurance for discriminatory harms is "efficient" depends upon the answer to the real-life question of who will ultimately end up paying the compensation awarded against employers found liable for discrimination. Insurance theory is useful in focusing attention on this issue. The economic analysis of products liability, for example, assumes that consumers will end up paying for the third party insurance mandated by liability for product defects because manufacturers will "pass through" the expense of compensation for consumer injuries in the form of higher prices. See *supra* text accompanying notes 153-55 for discussion on pass-throughs and cost-shifting.

In employment discrimination, it is unclear whether and to what extent employers will "pass through" the costs of paying liability judgments to the persons (e.g., minorities and women) who are eligible to collect, or whether those costs will be shifted to others (e.g., stockholders, product consumers, clients, customers, or other employees). For example, employers might avoid employing minorities or pay them less. Or, they might respond to higher labor costs by employing fewer workers generally or by substituting away from labor to capital-intensive activities. *See supra* text accompanying notes 151-52. Alternatively, they might raise prices to consumers, or simply absorb the costs themselves, as reflected in lower profits for the enterprise or less gain for stockholders. *See, e.g.*, Kennedy, *supra* note 153, at 604-07 (discussing distributional issues raised by various tort liability schemes, and raising questions regarding the "pass through" assumption that is central to scholarship in the economics of products liability).

Because any one of these scenarios is possible, any analysis—like the insurance theory of efficient damages—that rests on the assumption that liability costs will ultimately be "passed through" to victims must be approached with caution.

226. But see *infra* Part II.D.3.a for additional discussion of this issue from the vantage of creating a rational compensation system.

227. *See, e.g.*, KENNETH S. ABRAHAM, DISTRIBUTING RISK 24-25 (1986) ("[T]he use of insurance as a wealth redistribution device is subject to many difficulties.").

insurance system should charge persons with insuring themselves against the risks of harms that result from their own immutable characteristics. In other words, risk pooling arrangements should take no account of involuntary traits (such as race and sex) that correlate with risk. On this view, individuals from groups most likely to suffer discrimination should not be asked to insure themselves; rather, the system should be completely blind to the factors that give rise to the special risk, and the costs of insuring those groups against harm should be spread more widely.[228]

2. Compensation for Unconscious Bias: All-Or-Nothing Liability

The analysis up to now has offered no compelling rationale for distinguishing between conscious and unconscious forms of bias in the provision of insurance or compensation to victims. But the analysis has proceeded as if fairly accurate compensation and actuarially sound insurance could be achieved. It has ignored factors that contribute to errors in the assignment of compensation. Such errors are important in judging the desirability of a compensation scheme from the point of view of compensatory justice. A system that routinely misdirects compensation cannot be justified on the ground that it makes its victims whole or restores them to their rightful position.

For liability systems generally, the best decision rule is one that "imposes liability entirely on the party who would indeed be liable under the governing substantive law if only all the facts could be known with certainty."[229] But the facts never can be known with certainty. Although it is no great insight to point out that all liability systems produce errors in compensation, the degree to which different liability schemes directed at different harms can avoid mistakes in identifying and compensating victims is a subject that has received less attention than it deserves.[230] We have already discussed how the potentially limited ability of judicial fact-finders to distinguish between "genuine" precautions against inadvertent bias and "pseudo-precautions" might undermine the deterrent potential of liability rules. In a similar vein, an analysis of the potential for errors in assigning responsibility that is built into different antidiscrimination regimes is important to an understanding of how well those schemes advance compensatory goals.

The ensuing discussion examines the type and amount of error that might be expected to occur in a system directed at compensating individual workers for unconscious disparate treatment. It concludes that the compensation errors generated by such a system are likely to be quite significant. If, as already suggested,[231] unconscious disparate treatment is "subtle"—either because it is

---

228. *See id.* at 26-29 (discussing normative "egalitarian insurance principles," which would "render such immutable characteristics as age and sex morally irrelevant to the appropriate distribution of risk," including who should bear the cost of insuring against risk).

229. David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 AM. B. FOUND. RES. J. 487, 496.

230. *See, e.g.*, Geistfeld, *supra* note 223, at 637 & n.85 (noting paucity of attention to errors in assigning liability in discussions of whether liability schemes achieve aims of compensatory justice).

231. *See supra* Part I.C.

uncommon, or unpredictably sporadic, or "shallow" (in that it "makes a difference" only in scattered cases)—many instances of discrimination may go undetected. Victims will be unable to meet the standard of proof and will remain uncompensated. But if the standard of proof is set lower to make up for these limitations, undesirable errors will be made in the opposite direction. Regardless of the standard of proof, the difficulties inherent in distinguishing true unconscious bias from innocent conduct in the employment setting can be expected to produce frequent mistakes in one direction or the other. There will either be too much or too little compensation for claimants as individuals and in the aggregate, and firms will be undercharged or overcharged for their actionable harms.

Most liability systems adopt an all-or-nothing recovery rule: Compensation is awarded if the plaintiff proves causation and other elements of liability by a designated standard of proof, which in civil actions is a preponderance of the evidence. The current liability scheme for workplace disparate treatment conforms to this all-or-nothing pattern. The key question for the judge or jury in straightforward "single motive" or "pretext" cases is whether a decisionmaker's consideration of the race or sex of the victim produced an adverse outcome—that is, whether the trait was the "but for" cause of some loss to the plaintiff. The plaintiff must ordinarily show discrimination by a preponderance of the evidence—that is, that more likely than not the action taken against him was due to a protected characteristic.

As explained above, claims of unconscious disparate treatment will quite often stand or fall on statistical evidence.[232] The historical practice under Title VII (with its seminal emphasis on hiring cases), as well as the practical difficulties of generating persuasive numbers, have influenced the statistical methods that plaintiffs have offered and courts have accepted in the discrimination context. The dominant methodology is "hypothesis testing," which seeks to determine with some degree of "significance" or "confidence" the probability that observed disparities between outcomes for protected and unprotected groups (for example, differences in hiring rates) could happen by chance.[233] This method and the inferences that follow from it are not tied in any obvious way to the conventional requirement that a plaintiff meet a designated standard of proof. Specifically, it is unclear how these techniques would be related to a finding that discrimination is "more likely than not" the cause of observed patterns.[234]

In contrast, the question of causation in toxic torts cases is often analyzed quite differently by, for example, comparing the background risk of contracting a disease

---

232. *See supra* Part II.C.2.

233. *See, e.g.*, Marcel C. Garaud, Comment, *Legal Standards and Statistical Proof in Title VII Litigation: In Search of a Coherent Disparate Impact Model*, 139 U. PA. L. REV. 455, 479 (1990); Paul Meier et al., *What Happened in* Hazelwood*: Statistics, Employment Discrimination, and the 80% Rule*, 1984 AM. B. FOUND. RES. J. 139, 145-52. For a persuasive account of problems with the courts' use of hypothesis testing models in the discrimination context, see Browne, *supra* note 126, at 495-96.

234. *See, e.g.*, Browne, *supra* note 126, at 496 ("[T]here is no simple relationship between the significance level and the burden of persuasion."); *see also* Garaud, *supra* note 233, at 467-68 (discussing uncertainties surrounding implications of hypothesis testing models for plaintiff's actual burden of proof).

with the incidence among workers allegedly exposed to a disease-causing agent.[235] It is often accepted that the preponderance standard is satisfied if the risk among the exposed worker population is more than twice as high as among a comparable unexposed group—that is, if workers are more than twice as likely as a nonworker population with similar background risk to get the disease. In that case, one could say that it was more likely than not that any particular worker's cancer was caused by workplace exposure. Put another way, more than 50% of the risk of disease can be attributed to a workplace influence.[236]

A toxic tort type analysis as applied to discrimination on the job would correspond to a showing, for example, that employees from non-minority groups were promoted without apparent justification more than twice as often as otherwise comparable members of minority groups. Those numbers could be claimed to demonstrate disparate treatment by a preponderance of the evidence because they provide some basis for asserting that the failure of a particular minority or female worker to be promoted is more likely than not due to race or sex. Yet, for reasons that are probably both evidentiary and historical, the analysis in discrimination cases rarely proceeds along these lines.

Although the precise relationship in discrimination law between the evidentiary standard and different methods of statistical analysis is obscure,[237] one thing is clear: Employment discrimination practice, even where statistics play a central part, establishes an evidentiary threshold of *some* kind. That the threshold may vary from case to case and may sometimes fall well below the theoretically appropriate "more likely than not" mark complicates matters further, but does not alter the basic all-or-nothing nature of the liability determination under current law.[238]

---

235. Both toxic torts and discrimination cases share the problem of selecting an appropriate control group for fixing background risk. For example, the baseline risk of cancer for a particular group of workers might not match the risk for the population as a whole. Likewise, in the workplace example, arguments could be made that the non-minority population of workers is not the appropriate control for calculating background risk of nonpromotion. That argument amounts to the familiar assertion that there are hidden variables that account for differences in treatment between workers from different groups on the job. *See supra* notes 126-31 and *infra* note 247.

236. *See* Bert Black & David E. Lilienfeld, *Epidemiological Proof in Toxic Tort Litigation*, 52 FORDHAM L. REV. 732, 767 (1984); Rosenberg, *supra* note 67, at 857-59. *But see* Mark Kelman, *The Necessary Myth of Objective Causation Judgments in Liberal Political Theory*, 63 CHI.-KENT L. REV. 579, 620 (1987) (criticizing the "greater than 50% risk" approach to the preponderance standard in workplace harm cases as arbitrary and as conflating causation and valuation); Joseph H. King, *Causation, Valuation, and Chance in Personal Injury Torts Involving Preexisting Conditions and Future Consequences*, 90 YALE L.J. 1353, 1374 (1981) (same). *Cf.* Herskovits v. Group Health Coop., 664 P. 2d 474, 479 (Wash. 1983) (allowing a patient the opportunity to recover for malpractice by showing that negligence reduced his chances of survival by less than 50%, but noting that "loss of chance . . . does not necessitate a total recovery" for all damages cause by death.).

237. *See supra* text accompanying notes 233-34.

238. One important complication is that the burden of proving that bias is determinative (or, more accurately, nondeterminative) sometimes shifts. In "mixed motive" cases under Title VII, the employer now bears the burden on the question of whether bias "made a difference" for the purpose of awarding compensation. If the employer persuades the fact-finder that the "same decision" would have issued even absent discrimination, the plaintiff is entitled to no compensatory or equitable relief. The implications of mixed motive practice for compensation are discussed *infra*

How does the all-or-nothing rule, in combination with the controlling threshold standard of proof, create the potential for compensation errors in the context of unconscious bias? If race or sex is found to be the "determinative" cause of an adverse employment action, then the victim obtains a full "make whole" measure of relief. If that finding is not made, the claimant gets nothing.[239] If a claimant who was in fact discriminated against is unable to meet the effective threshold evidentiary requirement (let us say 50%) because, in effect, he is unable to prove that there is more that a 50% chance that *he* was the victim of "but for" discrimination, there will be an error in the payment of compensation against him and in favor of the employer. If he is able to meet the standard of proof, even though he was not in fact the victim of "but for" discrimination, the error in compensation will cut in the opposite direction.

What determines how often "true" victims of unconscious bias will be unable to prove their case, or how often nonvictims will be able to do so? The answer requires returning to the insight that unconscious discrimination, by its nature, must be proved statistically—that is, through analysis of multiple instances or examples of similar conduct. To be sure, statistical proof does play some role in cases of discrimination based on deliberate animus. Although people who act "on purpose" may know their own minds, others cannot see into them. Proof of conscious bias therefore also requires attention to outcomes. But although the proof needed to demonstrate unconscious bias will not necessarily differ qualitatively from that used to prove purposeful discrimination, it may well differ quantitatively. If unconscious bias is sporadic, unpredictable, and frequently nondeterminative, whereas conscious bias is more often predictable, determinative, and consistent across cases, then the assignment of compensation for unconscious bias will more frequently be in error under the all-or-nothing rule. Specifically, as the subsequent discussion shows, if unconscious bias only infrequently serves as the "but for" cause of unfavorable decisions against minority workers, then errors in compensation for unconscious disparate treatment will be large and recurring.

a. The Recurring Miss, or Lost Chance, Scenario

Suppose that the group-based biases harbored by a firm's supervisors affect the outcome of promotion decisions for only one in ten of all minority employees. The evidentiary "trace" left by the operation of these biases will, at best, look something like this: The numbers will show that minorities are promoted at a rate that is 90% of the rate among similarly qualified non-minority employees. (This hypothetical assumes what is rarely the case: that all employees in the comparison groups match on all attributes known to be job-relevant. If employees are dissimilar, plaintiffs would have to resort to regression analysis to demonstrate unexplained group

---

note 261.

239. The same applies to members of groups claiming disparate treatment, either in agency pattern and practice lawsuits or in private class actions. If an employer is found to have engaged in a discriminatory practice and that practice affects the fate of a group of employees, all victims get full relief. Otherwise they all come away empty-handed. *See* LINDEMANN & GROSSMAN, *supra* note 20, at 1777-79; Munroe, *supra* note 57, 228-30.

disparities in promotion rates.) Depending on the actual number of promotions and employees in each group, this will generate a rate of nonpromotion among minority employees that is only marginally greater than the rate among comparable employees from other groups.[240] What the data will not reveal—nor can it—is which *specific* persons within that group have actually suffered disparate treatment. Thus, even if the standard of proof is set extremely low and the statistical evidence offered by an individual plaintiff is deemed to meet the standard of proof, it would be impossible to know if that plaintiff actually deserved compensation. In fact, there would be only a one in ten chance that the compensation he received was in fact warranted by the "true facts."

On the terms of the hypothetical, a plaintiff (or group of plaintiffs) can only hope to show that there is a 10% chance that bias caused any one individual to miss a promotion. This is the same as showing that the fate of one in ten employees was affected by discrimination. But suppose that the law requires a showing of more than a 50% chance that discrimination was the cause.[241] Then any plaintiff who was in fact discharged because of actionable discrimination will receive nothing and will be undercompensated by the full amount of the damage he suffered. Since the best numbers any individual employee in the group can generate are essentially the same, no employee will be able to win his or her case. The firm will be undercharged overall for the harm it has caused, and the employees will be undercompensated as a whole, too. However, not every individual employee will be undercompensated, and thus the compensation error will not be evenly distributed among the employees in the group. Only those minority employees who were actually harmed (10%) will be undercompensated. Assuming no other significant source of fact-finding error, the rest will receive what they "deserve."[242]

---

240. If 20 out of 100 nonminority employees are promoted, then only 18 out of 100 similar minority employees will be promoted (because one in ten—or two in twenty—will fail to be promoted due to bias). So 80 nonminorities will fail to be promoted (an 80% chance of nonpromotion), and 82 minorities will fail (an 82% nonpromotion rate). The rate of nonpromotion among minorities will be 2.5% higher than among nonminorities ((82% - 80%)/80%). If 30 out of 100 nonminorities are promoted, then 27 minorities will be promoted. The failure rate among minorities will be about 4% greater than among nonminorities ((73% - 70%)/70% = 4%).

241. Once again, it is anybody's guess whether the courts do actually require a showing that corresponds to more than a 50% chance of causation. *See supra* text accompanying notes 233-34. And it is not clear what courts would in practice require if claims based expressly on accusations of unconscious bias became more accepted and routine.

242. *See* Neil Orloff & Jery Stedinger, *A Framework for Evaluating the Preponderance-of-the-Evidence Standard*, 131 U. PA. L. REV. 1159 *passim* (1983) (providing a description of pattern, magnitude, and distribution of errors expected under an all-or-nothing recovery scheme applying a preponderance standard).

The discussion assumes that the plaintiff is made to carry at the least an initial burden as to causation. Even in mixed motive cases, plaintiffs must show that discrimination was a "motivating factor." Civil Rights Act of 1964 § 703(m), 42 U.S.C. § 2000e-2 (1994). That requirement almost certainly permits plaintiffs to demonstrate something less than a 50% probability that discrimination was the "but for" cause of an unfavorable outcome. Meeting that standard, however, will not always entitle a plaintiff to compensation or equitable relief. See *infra* note 259 for a discussion of compensation in mixed motive cases.

If unconscious bias is uncommon, or if it is pervasive but "shallow" (i.e., usually nondeterminative), then the chance that it will "make a difference" to the fate of any particular employee could well be less than 50%. Even assuming that the weakest cases never get filed, employees suffering setbacks at work will often fail to prove that unconscious bias was "more likely than not" the cause of their injury. This scenario corresponds to the so-called "recurring miss" or lost chance situation,[243] in which victims rarely win or receive compensation (and tortfeasors rarely lose or pay) for uncommon harms. Although the large fact-finder errors that can be expected in real-life discrimination suits will allow many plaintiffs to win against the odds, others may lose often enough to generate significant undercompensation overall (which will undercharge employers for the harms generated).

One way to attempt to address this scenario of underrecovery is to set the threshold of recovery very low. Some features of current law appear to represent an attempt to do just that: For example, the requirement in mixed motive cases that plaintiffs prove only that group status was a "motivating factor" seems designed to boost the chance liability will be found even when the available evidence supports only a small probability that discrimination was the "but for" cause of an individual's loss of job benefits. But setting the all-or-nothing threshold low creates the risk of generating serious error in the opposite direction. Where there is only a small chance (say 10%) that bias is the cause of an adverse decision—which, in the case of unconscious bias, would produce a marginally higher incidence of unfavorable job outcomes for similarly situated members of one group—then fully nine out of ten of adversely affected minority employees would not in fact be the victims of determinative bias. Nevertheless, all of them will be able to meet the threshold standard. If all can recover in full, 90% of plaintiffs will receive a windfall. An employer who is sued repeatedly will seriously overpay. Thus, lowering the evidentiary standard cannot be a satisfactory solution to the problem of compensation errors generated by a pattern of recurring misses, since that move will give rise to large and repetitive errors in the opposite direction.

As this analysis reveals, the mere existence of errors in compensation in an all-or-nothing recovery scheme is not *solely* a function of the low probability of the actionable event, nor of the decision to apply a lax evidentiary standard. Rather, recovery errors will, by definition, occur whenever the actuarial probability that an actionable event is the "true" cause of harm falls short of 100%. Compensation errors will favor plaintiffs if the actual (and demonstrable) probability falls above the evidentiary threshold, and will favor defendants if it falls below.[244] Although the very existence of error in recovery is not dependent on the features here specifically

---

243. *See* Kaye, *supra* note 229, at 514 n.76; Saul Levmore, *Probabilistic Recoveries, Restitution, and Recurring Wrongs*, 19 J. LEGAL STUD. 691, 705-10 (1990); Robinson, *supra* note 129, at 792-93; Robinson & Abraham, *supra* note 188, at 1484-90; Rosenberg, *supra* note 67, at 877-79.

244. If an employee can show, for example, a 90% chance that actionable bias caused his firing, he will recover under a preponderance standard regardless of whether he was harmed or not (and there is a one-in-ten chance he was not). Every one of his (similarly situated) fellow employees will also recover, which will wind up overcharging the firm in the aggregate for a harm for which it was only 90% responsible overall. Moreover, some employees (the 10% who were not harmed) will receive a windfall.

attributed to unconscious discrimination—intermittent or infrequent determination of harmful outcomes—those features have a very important effect on the expected direction and magnitude of error in an all-or-nothing system. On a conventional preponderance standard (which may or may not be the standard under current antidiscrimination law), very few "true" victims will be compensated. Defendants who are repeat players will also be undercharged. But lowering the evidentiary standard to address this problem means that *many* nonvictims will be compensated and the employer will be seriously overcharged. (This aspect of inaccuracy will grow worse as events become less probable.) An event that is relatively uncommon—as unconscious disparate treatment may be—generates the unpalatable choice, in an all-or-nothing system, between nonrecovery for victims coupled with significant undercharging of firms, or large windfalls for nonvictims coupled with significant overcharging of firms.

Which scenario—undercompensation or overcompensation—would more likely prevail if liability were expressly extended to unconscious disparate treatment? If cases were analyzed on the "single motive" model, the existing framework might generate a "recurring miss" situation that threatens to undercompensate victims—although significant variation due to errors in fact-finding could conceivably blunt this effect. Alternatively, if unconscious bias cases were treated as mixed motive cases (as arguably they should be),[245] then the patterns generated in litigation would depend initially on how easy it would be for plaintiffs to make their threshold causal showing of an illicit "motivating factor." This would in turn depend on whether courts were willing to accept simple and small disparities in outcome by race or sex as satisfying the plaintiff's evidentiary requirement, or whether a more rigorous statistical analysis or other kinds of evidence would be required.[246] Recovery would also depend on how hard it would be for defendant firms to convince triers of fact that the "same decision" would have been made absent unconscious bias—a task that would depend critically on the niceties of statistical analysis and employers' ability to persuade fact-finders that benign factors explain differences in outcomes.[247] If the employer fails at this task, many plaintiffs might recover undeservedly, and firms might end up paying too much. In short, it is not clear whether extending the all-or-nothing liability framework to unconscious discrimination claims would result in a pattern of chronic overrecovery for victims

245. *See supra* text accompanying notes 49-53.

246. Although courts now generally require some type of particularized or anecdotal evidence to trigger a mixed motive analysis, retaining this requirement for unconscious bias cases would make little sense since the overtly biased attitudes required to generate such evidence would often be lacking. See *supra* text accompanying note 58 for a discussion of "direct evidence" requirements.

247. *See supra* text accompanying notes 125-32. Although there is no legal requirement as such that employers show that the factors relied on are related to productivity or are otherwise "rational," the fact-finder may give more weight to such a showing. See *infra* text accompanying note 251 on the assumption that employers act rationally, and *supra* text accompanying notes 130-31 on difficulties in demonstrating links to productivity. But see *infra* Part II.D.2.b for the argument that expressly extending Title VII to unconscious bias may weaken the presumption that employers always have good reasons for their decisions.

(and overcharging of firms) or underrecovery (and undercharging). But plausible scenarios suggest that errors would be large in either direction.[248]

b. Proving Irrationality

In addition to the difficulties already discussed, there is yet another factor that could influence the magnitude and pattern of compensation for unconscious bias and potentially exacerbate any tendency towards errors in recovery for inadvertent, as compared to more purposeful, forms of discrimination. This factor relates to the practical realities of litigating claims of unconscious discrimination. If employment discrimination theory or practice were changed to permit claimants to frame their claims openly as accusations of inadvertent disparate treatment, that would almost certainly lead to further weakening or abandonment of the *McDonnell Douglas* formulation, at least where claims of unconscious bias were concerned.[249] But whether encouraging express allegations that unconscious racial or sexual bias is at work would make discrimination harder or easier to prove could well depend on the *psychological* consequences of introducing the idea of inadvertent stereotyping to triers of fact. Exposing fact-finders to the notion that bias can operate unconsciously might make them more or less reluctant to infer that *trait-based discrimination* is the explanation for otherwise unexplained disparities in an employer's treatment of persons from different groups.

On the one hand, juries may be more reluctant to infer actionable bias from unexplained patterns because the idea of unconscious bias may suggest that employers are subject to unconscious motivations generally. And such motives, because hidden, involuntarily, and the product of poorly understood forces, need not be rational. As the Supreme Court's decision in *St. Mary's Honor Center v. Hicks*

248. One additional factor that might lead to overcompensation, however, is that inviting express claims of unconscious bias might encourage more disparate treatment claims to be filed on the basis of bare disparities in group outcomes alone, since plaintiffs can argue that the absence of anecdotal evidence of conscious animus does not undermine their case. If employers frequently make use of neutral criteria with a disparate impact, an increase in the number of these claims could give rise to more "false positives," or erroneous jury awards in favor of plaintiffs alleging unconscious bias. See *supra* note 131 on disparate treatment masquerading as disparate impact and vice versa.

249. As discussed above, *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973), does not rule out the possibility of liability where motives are unconscious to the extent that it permits plaintiffs to show disparities in treatment with no ostensible justification. *See supra* pp. 1151-52. Also, the *McDonnell Douglas* requirement that an employer supply reasons simply mandates what most defendants will do anyway, even if the allegation is one of unconscious bias. Nevertheless, it is hard in unconscious bias cases to justify the retention of the *McDonnell Douglas* requirement that the defendant supply reasons which the plaintiff must then prove "pretextual" if "pretext" depends on a defendant's false report of what is known to him. Where decisions may be influenced by unconscious forces, a finding that the reasons provided were "truthful" (in the sense of sincere, rather than in the sense of providing an accurate picture of the mind's unconscious workings) would not establish an absence of discrimination (and no liability). Nor would a finding that the defendant's reasons were false (in the sense that he was lying about them) even be probative of unconscious discrimination (or liability). The question of whether unconscious bias played a role in the decisionmaking process would be completely open in either case.

made clear, mere disparities, even in the absence of an explanation for those disparities, does not necessarily mean that forbidden discrimination is at work.[250] But neither does it mean that the employer had a "good reason" for what he did. Rather, some research in cognitive psychology—research to which juries may well be exposed as plaintiffs try to convince them of the reality of "unconscious bias"—is notable in showing that human judgment often falls short of perfect rationality.[251] Thus, in addition to turning the defendant's sincerity or insincerity into a side-show, the emphasis on unconscious bias as a factor in workplace decisionmaking suggests to the jury that supervisors do not always act for a good reason or for reasons that advance the employer's interests. If employers cannot help but apply irrational group-based categories, then perhaps they are at the mercy of other cognitive imperfections as well. To take this possibility seriously is to call into question two of the implicit assumptions that are vital to the proof of employment discrimination cases under current law: that employers who do not discriminate (that is, are not motivated by race) are motivated by productivity-related criteria and, conversely, that employers who are not motivated by criteria demonstrably related to their own best interests must be motivated by impermissible factors such as race.[252] The very notion that employers are influenced by unconscious "mental habits" or uncontrolled, irrational biases, because it suggests that not all unexplained actions can be attributed either to discrimination or to the employer's self-conscious pursuit of profits, may make the fact-finder more skeptical of the claim that *actionable* unconscious bias is at work in any given case. This might make it easier for employers in mixed motive cases, for example, to rebut the presumption that an illicit "motivating factor" was the "but for" cause of an employment decision. The

---

250. 509 U.S. 502, 506-512 (1993).

251. For an extensive discussion and review, see Robyn M. Dawes, *Behavioral Decisionmaking and Judgment*, in 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY, *supra* note 16; Wilson & Brekke, *supra* note 11, at 118-19, 126-30. *See also, e.g.*, Malamud, *supra* note 39, at 2254-58 (describing arbitrators' characterization of much decisionmaking in the employment context as irrational or arbitrary even though not animus-based).

Although, as discussed *supra* Part I.D, unconscious biases may at times be "rational" in taking advantage of valid group-based generalizations, stress on the rationality of some group-based biases is not likely to loom large in the litigation strategy of plaintiffs in unconscious bias cases.

252. *See supra* note 53; *see also, e.g.*, Laycock, *supra* note 126. As Krieger states:

> Pretext analysis permits this inferential leap from an apparently irrational or inconsistent judgmental process to an intentionally discriminatory one through the operation of a "presumption of invidiousness" first articulated by the Supreme Court in *Furnco Construction Corp. v. Waters*....
>
> ....
>
> Pretext analysis thus rests on the assumption that, absent discriminatory animus, employment decisionmakers are rational actors. They make evenhanded decisions using optimal inferential strategies in which all relevant behavioral events are identified and weighted to account for transient situational factors beyond the employee's control.... The presumption of invidiousness permits the trier of fact to infer discriminatory intent from flaws in a decisionmaker's inferential process. Without this presumption, one could only infer that an irrational decision was made; such a decision, in the absence of a duty to discharge only for good cause, would not be actionable.

Krieger II, *supra* note 1, at 1181.

result would be a reduction in plaintiffs' chances of recovery overall with even less recovery for deserving victims.

On the other hand, exposure to scientific evidence for unconscious categorization or trait-based stereotyping might lead triers of fact to believe that inadvertent bias against disfavored groups is a pervasive and constitutive feature of workplace life. Information about unconsciously biased patterns of thought, by creating the impression that stereotyping is a widespread and unavoidable mental habit, could have the effect of "normalizing" discrimination and even destigmatizing its practice. Inadvertent discrimination would understandably be regarded as less morally blameworthy than animus-based bias. This might make triers of fact less reluctant to find that employers have discriminated despite good faith disavowals of prejudicial intent. The result would be more frequent victories for plaintiffs claiming unconscious discrimination, with more money undeservedly flowing to nonvictims.

3. Compensation for Unconscious Bias: Probabilistic Recovery

This Part addresses an alternative approach to an all-or-nothing recovery system: A probabilistic system that gears the *amount* a claimant can recover to the probability that a harm was due to an actionable cause. It concludes that probabilistic recovery does not represent a viable alternative to an all-or-nothing recovery rule for unconscious disparate treatment.

Under all-or-nothing recovery, each victorious plaintiff receives compensation for the full value of his claimed compensable loss. Under a probabilistic scheme, each plaintiff is awarded an amount proportional to the calculated expected contribution of the actionable cause—which would here be racial or sexual cognitive bias—to the decision in his case. "The proportionality rule discounts recovery by the probability that the plaintiff's loss was caused by some other wrongdoer, by a nonculpable source, or by the plaintiff."[253] In the biased supervisor example, an individual plaintiff would, at best, be able to show that ten ostensibly comparable nonminority employees were promoted for every nine employees like himself. But that evidence provides the basis for, at most, a probabilistic statement: that there is a 10% chance that any member of a group to which he belongs was, in true fact, the victim of disparate treatment. Under a probabilistic recovery rule, the plaintiff would be entitled to 10% of the full measure of compensation for losses estimated to be due to the unfavorable employment outcome alleged (failure to promote). The same analysis would apply to any group of plaintiffs suing together rather than seriatim. Each would recover 10% of full compensation for losses from the adverse event.

To be sure, a probabilistic recovery or "expected value" rule—unlike the all-or-nothing rule—"errs in every case."[254] In those individual cases in which bias is in true fact the "but for" cause of the harm, the rule undercompensates. Where bias is not in fact the "but for" cause, it overcompensates. But where statistical evidence reflects actual frequency, it operates to award the correct amount of compensation

253. Rosenberg, *supra* note 67, at 881.

254. Kaye, *supra* note 229, at 502.

to the "at risk" group as a whole. Moreover, unlike the all-or-nothing rule, the probabilistic rule not only awards the claimants but also charges the perpetrator with the right amount in the aggregate.[255] The rule has the virtue, on the deterrence side, of not mandating any systematic "overcharge" of the firm—or, for that matter, any undercharge—for the cost of its tortious activity over the long haul. (Unfortunately, as explained more fully below, that virtue is worth little in the unconscious bias context: there is no reason to predict that, even under a probabilistic recovery rule, damages will equal harm and deterrence will be efficient.[256])

a. Strengths of Probabilistic Recovery

From the point of view of victim compensation, would a probabilistic recovery scheme for unconscious bias be more desirable than a rule that more closely tracks the structure of current law? That depends on whether an all-or-nothing rule would most likely generate overrecovery or underrecovery—an empirical question quite difficult to answer in the abstract.[257] Which pattern would prevail—and whether probabilistic recovery would represent an improvement—would depend on unknown facts about the "true" incidence of unconscious bias as well as on whether courts would tend to treat unconscious bias claims as ordinary "pretext" claims or as "mixed motive" claims with burden shifting rules that resemble those now in place. As already discussed, placing the entire burden of persuasion on claimants might generate a "recurring miss" or "lost chance" scenario: assuming a low incidence of determinative bias, most victims would not recover. In that case, the choice on the compensation side would be between an all-or-nothing scheme that produces recovery for very few "true" victims, and a probabilistic scheme of prorated recovery that produces modest windfalls (for nonvictims) but also some partial recovery for the deserving. If virtually total nonrecovery for "true" victims (false negatives) is seen as worse than the combination of partial recovery for victims plus some recovery for nonvictims (partial false positives), the expected value rule seems clearly superior. Alternatively, if a "shifting burden" or mixed motive framework were applied, and that framework resulted in overrecovery in an all-or-nothing regime, probabilistic recovery would more effectively avoid overcharging the discriminatory enterprise and might also reduce the amount of "windfall" for many nonvictims. However, it would also diminish the amount of compensation that would otherwise be awarded "true victims"—a more equivocal result.

But apart from potentially increasing the frequency of victim recovery, the probabilistic approach has other important features that argue in its favor: The rule is truer to the phenomenon of unconscious disparate treatment in the workplace and points clearly to the nature of the inquiry that must be undertaken when investigating unconscious bias. Given the evidentiary limitations inherent in trying to observe the workings of other minds, unconscious bias is best viewed as presenting an actuarial

255. *See id.*; Rosenberg, *supra* note 67, at 885 ("The defendant never overpays, and the population as a whole gains no windfall.").

256. *See infra* Part II.D.3.b.

257. *See supra* text accompanying notes 245-48.

risk of harm to an indeterminate plaintiff. By observing outcomes, it is at best possible to assign a frequency to the harmful event. It is not ordinarily possible, however, to identify specific victims, either ex ante or ex post. To the extent that a probabilistic approach honors these limitations, it will lead to greater acceptance and understanding of the uncertainties surrounding the detection of unconscious bias in the workplace setting.

Another benefit of adopting a probabilistic approach to unconscious disparate treatment claims is that it would implicitly resolve the question of when plaintiffs in mixed motive cases should be allowed to recover. As explained, the mixed motive paradigm as applied to typical cases of subjective unconscious bias provides a way of getting at whether bias-in-the-process is in fact determinative of outcomes. The question of whether bias in the process—that is, bias as a "motivating" but not the "but for" factor in an employment decision—is itself an actionable harm (although it produces no material injury) can be viewed as the unfortunate product of a misleading reification of a fundamentally probabilistic concept. As the discussion thus far suggests, the concept of bias as a "motivating factor" is better thought of as a stand-in for the risk that an illegitimate factor, such as race or sex, will distort a decision enough to issue in a tangible harm. If an "expected value" recovery rule is adopted, many plaintiffs (in keeping with the recommendations of some scholars[258]), will receive some compensation for exposure to biased decisionmaking regardless of whether they are actually harmed by it. That recovery is better conceptualized as a form of "compensation for risk," rather than compensation for some form of intangible or "emotional" injury.[259] And the right to recover is best viewed as a practical compromise, constructed in the face of ineluctable uncertainty, that tolerates some errors in assignment and amount of compensation to individuals for the sake of a potentially more accurate assessment of costs to the tortfeasor

---

258. *See supra* note 222.

259. The observation that compensation to individuals who are exposed to the risk of harm, regardless of whether that risk is determinative in their case, is not really a form of compensation for *distinct injuries* inflicted by exposure to risk is consistent with the refusal to give compensation to persons who are exposed to risk but never suffer any harm at all.

To illustrate this point, suppose that a person is evaluated in a manner tainted by an impermissible consideration such as race, but no adverse action is taken against him. That person would have no colorable claim even for mixed motive recovery, since he suffered no adverse setback in interest at all. Put another way, he never suffered the injury of which exposure to bias enhances the risk. Yet a person who was evaluated in a biased manner, but who was fired or otherwise unfavorably treated for *other reasons*, would be entitled to recover. Both persons were exposed to an enhanced "risk" of harm from the defendant's actionable conduct. The only difference between them is that the second person happened to suffer injury due to another (nontortious) cause and the first did not. Yet they would be treated differently under a probabilistic regime. This difference in treatment makes no sense if recovery is seen as compensation for distinct injuries caused by the mere exposure to a "risk of harm" in and of itself. But that is not the right way to think about probabilistic recovery. Indeed, recovery for all exposure to risk would result in overcharging the defendant overall because, as a practical matter, it would overestimate the share of harm attributable to the defendant's agency. Rather, the point is to peg recovery to the actuarial probability that the losses that *did* occur were in fact due to the defendant's tortious behavior.

overall.[260] The compensation each member of the group receives is not necessarily for any "real" injury the person has suffered through such exposure (although it might be). Rather, compensation is geared to the actuarial possibility that each person's loss was in fact caused by the tortfeasor and not by another factor.[261]

A probabilistic rule also has the advantage of fitting with a notion of mixed motives that is not only more psychologically accurate but also truer to the epistemological limitations inherent in proving facts about other minds that are hidden even from those minds themselves. Mixed motive jurisprudence appears to assume that the employer has multiple *conscious* reasons for action. The idea of race as one "motive" among many might correspond to our subjective experience of making decisions that take many considerations into account,[262] but it provides a poor description of what is going on when unconscious bias is at work. Going back to the example of the supervisor whose unconscious biases cause him to disfavor 10% of his minority employees: in that case, the supervisor may report one or more reasons for his actions. Insofar as he is privy to his own conscious (as opposed to unconscious) states of mind, he will be telling the truth. But in fact his self-report is incomplete: He does not know that hidden cognitive mechanisms have distorted the application of his reasons in a few cases.

---

260. Since the point is also to achieve actuarially accurate compensation for tangible harms suffered by all victims as a group, the insurance component of a probabilistic recovery scheme for unconscious bias does not, despite appearances, build in any compensation for "pain and suffering" or emotional harm, and thus is not vulnerable to the kinds of objections that could be leveled against mandating insurance for these types of losses. See *supra* text accompanying notes 222-24 for a discussion of insurance for nonpecuniary harms.

261. At least one advocate of full compensation in all mixed motive cases has considered probabilistic recovery as a possible solution to the problem of nondeterminative causation, only to dismiss that approach summarily. *See* Weber, *supra* note 9, at 529. Weber's rejection of a probabilistic averaging rule is based on the fallacious assumption that, once a defendant is found to have "committed discrimination"—that is, once the plaintiff has met the applicable standard of proof—all uncertainty as to cause disappears. *Id.* But Weber confuses the case where causal probabilities are prospectively uncertain but become quite certain once the event occurs—as with the lightening strike example Weber provides—with a case like unconscious discrimination, in which the cause remains as obscure after the event as before. For unconscious discrimination, prospective and retrospective estimates of risk differ very little.

262. Certainly, where all our "reasons" appear transparent to us, we might subjectively believe we can gauge the extent to which different considerations actually influenced or played a part in the decision that is finally taken. We might, for example, believe that "this influenced me a little, while this other factor influenced me a lot." We might even venture to report that a particular factor did or did not satisfy the "but for" test for a decision because we are fairly confident that, if a particular factor were not in play, we would have made the same decision or we would not. But as suggested by the discussion on "source confusion," *supra* text accompanying notes 93-95, we often cannot claim to issue an accurate report on the anatomy of our own deliberations. In particular, we cannot possibly hope to know *in any specific case* how much each factor in the mix influenced us. We can only hope to know something that has little to do with the workings of any individual's mind: the probability over the general run of cases that a particular factor "made a difference."

That the extent or nature of our knowledge might differ dramatically when our motives are largely "conscious" (if there ever is such a case) as opposed to when unconscious factors dominate should come as no surprise. Introspection is a powerful avenue of access to valuable information. When critical deliberative processes are hidden from us, that avenue is closed.

This story represents one type of mixed motive situation that has been almost completely ignored in current law. The existing paradigm is geared to identifying the contribution of a set of discrete "motivating factors" to *each* decision. Yet it is not quite right to say that race made some precise contribution—let us say ten percent—to the supervisor's judgment of *each* employee. There is no evidentiary "trace" that can reveal the precise ratio of valid "motive" to suspect motive for any *particular* decision. The best we can do is estimate some probability that unconscious bias made a difference in any one case. This always requires looking at *more* than one case. The mix of factors that produced each individual decision simply cannot be known.

b. Drawbacks of Probabilistic Recovery

Although a probabilistic system has many strengths in theory, formidable practical obstacles make this system unworkable. A probabilistic rule that requires assigning a precise probability to the elements—including unconscious discrimination—that contribute to any workplace decision would strain the fact-finding capacity of a liability system to the breaking point.[263] The methods of statistical estimation and regression analysis that have been developed for use with an all-or-nothing system will almost certainly be incapable of supporting such an analysis. The challenges inherent in trying to come up with precise probability estimates without adequate evidence would run the risk of producing a plethora of errors. This would surely represent no improvement in accuracy over conventional all-or-nothing determinations.[264]

The focus in this Part has been on the goal of compensating victims completely and accurately for harms incurred. But the desirability of an expected value rule will also turn on its incentive effects. In theory, probabilistic recovery comes closer than an all-or-nothing recovery rule to charging enterprises for the true costs of their actionable harm when firms are "repeat players."[265] If the recurring miss scenario would dominate under all-or-nothing, firms would almost certainly pay more under

263. *See* Rosenberg, *supra* note 67, at 898 ("Under the preponderance rule, the trier of fact is concerned only with the question whether the probability exceeds fifty percent; the precise amount by which the probability falls above or below that threshold is irrelevant. Imposing proportional liability, on the other hand, would obviously require far more precision.").

264. Moreover, even setting aside evidentiary limitations, there would be additional problems related to relief. The remedies available under probabilistic schemes for group recovery in other settings are purely monetary. *See* Robinson, *supra* note 129, at 785-89, 794-95; Rosenberg, *supra* note 67, at 908-24. But an important component of relief for employment discrimination has always been equitable reinstatement, which is ideally suited to an all-or-nothing recovery rule and cannot easily be adapted to an expected value regime. In lieu of equitable relief, some measure of prospective losses would have to be incorporated into any remedial calculation. Arguably, monetary recovery could never be a perfect substitute for equitable reinstatement. The sacrifice of this component of relief would make such a compensatory scheme less desirable from the victim's point of view.

265. *See* Levmore, *supra* note 243, at 697-98 (stating that accurate internalization of costs from probabilistic recovery depends on defendants engaging in many rounds or instances of similar conduct).

an expected recovery rule. If the "shifting burdens" scenario is closer to reality, they might pay less.

These observations must be considered in light of the conclusions in the first part of this Article: having enterprises pay more is not necessarily better, even when "more" reflects the actual costs of harm inflicted. That is, requiring employers to pay even actuarially sound compensation could produce perverse effects by tempting employers to reduce activity levels and take wasteful "pseudo-precautions," or by shifting costs away from the cheapest cost avoiders. Thus, internalization of all costs of unconscious bias to the employer, even if it could be achieved, is not an unalloyed good. The virtues on the compensation side of more ample (and actuarially accurate) recovery must always be traded off against the undesirable effects of potentially greater liability for firms.

In the end, it must be understood that even if a probabilistic scheme for remedying unconscious bias could be made to work, it would still be no more accurate than an all-or-nothing rule in matching compensation to deserving victims. That is, it would repeatedly fail to insure that only true victims receive compensation and that nonvictims do not. That neither liability system can avoid significant inaccuracy in this respect suggests an important point: The patterns of compensation for unconscious bias can be expected to converge with the assignment of benefits and costs from programs that extend preferential treatment to members of protected groups. Such programs are sometimes criticized as poor vehicles for accomplishing any valid remedial purpose because the tie between compensable loss and conferred advantage is haphazard: There is no guarantee that persons helped by affirmative action have been victimized by discrimination in the past, and many victims of past discrimination will never benefit from any preferential treatment at all.[266] But that loose tie between victimization and reward will inevitably characterize any individualized liability system designed to address unconscious bias. The more subtle and erratic the bias, the more errors will be made in detecting it. The more errors made, the more often compensation will miss the mark. The expectation is that victims will frequently be rewarded and nonvictims will not. The pattern generated will bear no necessary relationship to any prior acts of discrimination. A liability scheme for unconscious bias will therefore engage an elaborate probative machinery and consume considerable judicial resources to achieve outcomes that mimic the results of diversity action or preferential treatment programs. That result, if desired, could be accomplished less wastefully by more direct methods.[267] Electing the more expensive and cumbersome route stands in need of justification.

---

266. This has led some commentators to disavow or downplay the importance of any remedial purpose for affirmative action. *See, e.g.*, Estlund, *supra* note 174, at 58-59; *see also* Susan Sturm & Lani Guinier, *The Future of Affirmative Action: Reclaiming the Innovative Ideal*, 84 CAL. L. REV. 953 (1996); Kathleen M. Sullivan, *Sins of Discrimination: Last Term's Affirmative Action Cases*, 100 HARV. L. REV. 78 (1986).

267. Compare *supra* Part II.C.5.b, arguing that, because liability for unconscious discrimination will cause employers to adopt diversity action programs (which might include a component of preferential treatment) as a way to reduce the risk of liability, its desirability should be assessed on the basis of that effect.

III. DISCRIMINATION AS ACCIDENT: WHAT IS TO BE DONE?

In the twelve years since Charles Lawrence published his seminal article on unconscious discrimination,[268] scholars have bemoaned the law's failure to effectively police and penalize unconscious forms of bias.[269] This Article has sought to analyze the sources of this putative inadequacy. It concludes that the most important obstacles to an effective remedy for unconscious disparate treatment lie not in some defect in the design or doctrinal framework of Title VII, but rather in the nature of the phenomenon of unconscious discrimination itself.

If the goal of a liability system is to diminish or compensate for the harms that it formally targets in a precise and cost-effective manner, then this Article suggests that a system directed at inadvertent bias will not accomplish those objectives very well. Liability cannot do its job without going off in search of the harm itself. But because we simply lack the tools to *detect* unconscious bias, we cannot directly *force* its elimination to the extent it does occur. That we cannot know another mind is a problem that plagues discrimination law generally. The dilemma is even more acute when the other mind can neither know itself nor effectively control itself nor be effectively controlled by others. The difficulties of detecting, measuring, and monitoring unconscious bias counsel avoidance of methods of attack that require precise determinations of causation or the accurate tracking of occurrence, magnitude, and effects. Moreover, the paucity of evidence that can reliably distinguish "true" bias from other causes, coupled with employers' complete control over employees' fate, may mislead fact-finders into believing that steps that employers take to guard against or compensate for unconscious bias work better than they do. If employers can thus reduce their expected amount of liability without proportionately reducing the costs of the targeted harm, they may expend more resources than are justified by an abatement of that harm or any other positive effects. Finally, employees may be better situated than employers to minimize biased assessments with the fewest social costs. Then imposing liability on employers could blunt victims' incentives to search for harm-reducing strategies or, at best, add little to existing incentives but at great cost in judicial resources. For all these reasons, there is no guarantee that liability will operate efficiently.

So what is to be done about unconscious disparate treatment? The answer depends in part on one's views about how a liability system is supposed to operate and on the legal system's proper role in effecting social change. This Article predicts that liability for unconscious disparate treatment will be inefficient and will fail to compensate victims accurately. One answer to this might be, "so what?" On this view, a legal response to race and sex discrimination is not ill-advised just because there is no guarantee of net social benefit or of individualized compensation. Perhaps antidiscrimination laws should not be about efficiency—that is, the balance of social costs overall—but about the fair distribution or just transfer of benefits. And perhaps the primary concern should not be with precisely compensating individuals at all, but rather with group entitlements and group fate. On this view,

---

268. Lawrence, *supra* note 1.

269. *See* sources *supra* note 1.

threatening liability for a previously neglected source of harm to vulnerable groups can be viewed as a blunt device for encouraging institutions to take an expansive array of measures with mainly prospective effects that appear to improve the lot of disfavored groups as a whole. More specifically, the main role of imposing liability for employment discrimination is to encourage employers to adopt "diversity action" type programs, regardless of whether such programs are "efficient" in light of the costs of the judicial machinery that administers the system or the overall effects of the programs themselves.[270] To be sure, proponents seem to hope that the reforms will alter at least some of the conditions that tend to foster inadvertent stereotyping. Alternatively, employers may change some practices that have a "disparate impact" on women's or minorities' progress. Or they might extend frankly preferential treatment to previously disfavored groups. They may do all of these things at once or in different measure. In the end, it does not matter exactly how the programs operate, as long as members of some groups do better, or appear to do better, within the organization.[271]

The desire to encourage broad-based, employer initiatives by any means possible is in keeping with a growing skepticism among legal scholars about the importance of the right to be free from disparate treatment as narrowly defined—a skepticism born of doubts about the effectiveness of the law's targeting of differential treatment "because of" race or sex. According to this view, disparate treatment—whether conscious or unconscious—is only a small part of what is holding minorities and women back in the workplace today.[272] Many problems now stem from ostensibly

---

270. Extending disparate treatment liability may also expand the opportunities for ordering these institutional changes directly or in the form of consent decrees. But this avenue will necessarily play a relatively minor role, since complex equitable relief of this scope is principally confined to class action or pattern and practice cases. Those cases represent a very small percentage of the claims that are presently brought under antidiscrimination laws, and it is hard to know whether this would change if unconscious bias were expressly covered.

271. One example of such a multi-pronged approach is a program adopted by the Department of Medicine at Johns Hopkins Medical School in response to a poor record of academic promotion and retention among women faculty. Among the steps taken by the department were: establishing a formal mentoring program that paired a female junior faculty member with a senior faculty sponsor, running regular informational meetings to offer concrete advice about career advancement, moving academic conferences and departmental meetings to ordinary business hours, conducting more careful and explicit periodic academic reviews, and encouraging senior male faculty members to invite junior female colleagues to conferences or to put their names forward as conference participants. Academic promotions among women faculty increased significantly during the five-year period the program was in effect. *See* VALIAN, *supra* note 2, at 319-20; *see also* Linda P. Fried et al., *Career Development for Women in Academic Medicine: Multiple Interventions in a Department of Medicine*, 276 JAMA 898 (1996) (describing the Johns Hopkins initiative).

272. Scholars tell stories of workplaces riddled with complex, "path-dependent" structures or practices that are kept in place by inertia, vested interests, coordination problems, transaction costs of switching to less discriminatory arrangements, lack of information about productivity, or the need to discourage shirking. "Pipeline problems"—the paucity of qualified or well-trained candidates for desirable jobs—also represent a potent source of differences in employment prospects. *See, e.g.*, Mary E. Becker, *Needed in the Nineties: Improved Individual and Structural Remedies for Racial and Sexual Disadvantages in Employment*, 79 GEO. L.J. 1659, 1661-63 (1991); Jomills Henry Braddock & James M. McPartland, *How Minorities Continue To Be*

neutral practices with disparate impact that would not remotely satisfy a "but for" test for disparate treatment by race or sex. Those obstacles persist under the current disparate impact framework, which allows businesses a great deal of leeway in defending neutral practices.[273]

In light of the view that the real impediments to the advancement of women and minorities in the workplace have so far eluded legal redress, it is tempting to argue that the law is too timid and narrow. Since more is better than less, extending the coverage of the liability system to reach additional harmful conduct, however expensive, must improve on the status quo. This view gains added momentum from the observation that judicially or legislatively compelled "diversity action" programs cannot hope to withstand scrutiny without fairly specific findings of past "sins of discrimination."[274] Within the limits established by our constitutional and remedial regime, the liability paradigm offers one of the few legally secure avenues for forcing broad-based changes in private workplace organization.

An uncritical positive outlook on any changes an employer might make in response to liability for unconscious discrimination defies the analysis in this Article, which takes seriously the traditional economic criteria for a rational liability scheme and refuses to give the benefit of the doubt to reforms of indeterminate value and unproven effect. Defending liability for unconscious bias not only discounts efficiency as an important goal, but also indulges the assumption that any changes made to avoid liability will represent an improvement of some kind for *someone*, either in the form of real reductions in unconscious bias, a general brightening of prospects for targeted groups, improvements in worker conditions generally, or increased productivity overall.[275] But once expected damages need not match

---

Excluded from Equal Employment Opportunities: Research on Labor Market and Institutional Barriers, J. SOC. ISSUES, Spring 1987, at 5, 6-24; Charny & Gulati, *supra* note 192, at 60; Daniel A. Farber, *The Outmoded Debate over Affirmative Action*, 82 CAL. L. REV. 893, 918-24 (1994); Samuel Issacharoff & Elyse Rosenblum, *Women and the Workplace: Accommodating the Demands of Pregnancy*, 94 COLUM. L. REV. 2154, 2159-71 (1994); Mark S. Kende, *Shattering the Glass Ceiling: A Legal Theory for Attacking Discrimination Against Women Partners*, 46 HASTINGS L.J. 17, 31-40 (1994); Strauss, *supra* note 29, at 1640-43.

Despairing of the inability of individualized liability systems to address the myriad structural and custom-based sources of female and minority disadvantage in the workplace, many commentators suggest structural solutions—including outright recourse to preferential treatment—that go beyond the formal limits of the remedies available under existing antidiscrimination laws. *See, e.g.*, VALIAN, *supra* note 2, at 319-32 (defending a variety of concrete "progressive policies" in the workplace as the best way to attack the myriad sources of differential outcomes); Donald L. Beschle, *"You've Got to be Carefully Taught": Justifying Affirmative Action After* Croson *and* Adarand, 74 N.C. L. REV. 1141, 1177-81 (1996) (arguing that affirmative action is necessary to offset intransigent racism); Brown et al., *supra* note 1, at 1490-92, 1528-30 (defending affirmative action as the only effective remedy for racism and racial disparities); Selmi, *supra* note 1, at 1296-1308 (defending affirmative action as the only effective remedy against discrimination).

273. For a review of the current law on disparate impact and the "business necessity" defense, see Linda Lye, *Title VII's Tangled Tale: The Erosion and Confusion of Disparate Impact and the Business Necessity Defense*, 19 BERKELEY J. EMP. & LAB. L. 315, 348-53 (1998).

274. *E.g.*, Sullivan, *supra* note 266, at 82-83.

275. *See, e.g.*, Gillian Flynn, *The Harsh Reality of Diversity Programs*, WORKFORCE, Dec. 1998, at 27; Gordon, *supra* note 161.

expected losses, there is no assured correspondence between the employer's response and a socially beneficial result. Indeed, employers' expenditures cannot be relied upon to provide minimal benefits of any kind, even for those groups the expenditures are designed to help, because there is no a priori basis for predicting whether expanding liability in this area will have any significant effect one way or the other. This Article proceeds on the assumptions that making unconscious bias actionable will strengthen the impetus to adopt "diversity action" programs beyond the incentives created by existing law, institutional priorities, or the march of cultural forces in general. But this may not be so. Then many of the detrimental effects described in this Article will not be forthcoming, but neither will the benefits posited by defenders of expanded coverage. All that will be left are transaction costs—that is, the costs of processing additional claims or of trying to prove unconscious bias. The point is that, in a world that uncouples harm from liability, it is hard to predict exactly what will happen. How employers will react and the effects of those reactions must ultimately be determined by empirical investigation.

If the law is not revised expressly to *extend* coverage to unconscious motivation, should it be changed in the opposite direction? Should Title VII be amended to state that recovery is allowed for purposeful bias only? The answer must start with the understanding, already discussed, that current practice almost certainly awards relief in some cases where unconscious bias is at work. Although *McDonnell Douglas* stands as an impediment to favorable decisions in many such cases, the methods of proof employed in some individual cases under existing law either do not permit fact-finders cleanly to separate deliberate from inadvertent motive or do not require them to do so. An important drawback of the present state of affairs is that it leaves employers in a state of great uncertainty. Employers may well believe that they are potentially liable for both deliberate and inadvertent bias and act accordingly.

It is unclear, however, whether rewriting the law to exclude inadvertent bias would significantly improve on this situation. Much of the argument in this Article hinges on contrasting the difficulties of detecting, monitoring, proving, and controlling disparate treatment that is inflicted consciously with that imposed unconsciously. But those differences must not be overstated. Although proving unconscious motive will rest almost exclusively on statistical evidence of unexplained differential outcomes, attempts to prove deliberate bias, even in individual cases, will sometimes make use of this type of evidence as well. Because the proof plaintiffs offer in both types of cases will often overlap, fact-finders might sometimes end up imposing liability for unconscious bias even if the law were expressly amended to permit recovery for purposeful bias only. It is thus hard to see how even eliminating the formal ambiguity in Title VII could reliably separate recovery for conscious and unconscious bias.

Moreover, forbidding recovery for inadvertent bias might also have the effect of creating a new type of argument that Title VII defendants could use to avoid liability. Perhaps because current law is ambiguous, employers are not in the habit of defending themselves against charges of discrimination by admitting race or sex-based differences in judgment but pleading inadvertence. Yet changing the law to distinguish sharply between conscious and unconscious causation would invite such a tactic by creating a "safe harbor" for unconsciously inflicted discrimination. By focusing attention on the conscious/unconscious distinction, this change might result in fact-finders demanding more persuasive evidence of the deliberateness of

defendants' discrimination than is now required. Plaintiffs' proof might fall short in significantly more cases and meritorious cases of purposeful discrimination could become harder to win.

Current law may well represent a "second-best" compromise between unattainable alternatives rendered problematic by uncertainty about how the system now works and how it would work if changed. As such, the argument for maintaining the status quo might go something like this: The system should focus on discriminatory behavior that can most cheaply and effectively be deterred or that permits the most precise assignment of compensation. Purposeful bias fits this bill as being most amenable to self-regulation, most responsive to the threat of sanction, and generally easiest to prove. As noted, current law may also catch within its net some examples of unconscious disparate treatment. Most likely these will include the most egregious cases of stereotyping, which generate the largest and most systematic unexplained differences in treatment. As argued, those cases will be atypical and thus uncommon. Although the price to be paid for this hybrid state is uncertainty about the law's coverage, which may spur employers to adopt an excessive number of self-protective measures, that cost is mitigated by the realization that, due to evidentiary overlap, some degree of imprecision in scope probably cannot be avoided even by the clearest statutory language. And leaving the law as it is avoids the possibility of making conscious bias significantly harder to prove.

Although this Article concludes that nothing more should be done within the existing legal framework to address unconscious disparate treatment, the conclusion is tentative. The issues raised in this Part would benefit from further analysis. Nonetheless, this stance should not be construed as denying that unconscious bias might play some role in impeding the progress of disfavored groups in the workplace or in society in general. It is simply impossible to know for sure how pervasive or important it is relative to other factors. Those who believe themselves victims of unconscious discrimination should recognize that unconscious bias is an unpredictable, mysterious, and elusive phenomenon that can only be tentatively inferred but never observed directly. They must be willing to acknowledge that suspicion is not fact, that "one is never one's own control group,"[276] and that the current state of understanding almost always impedes definite attributions of blame or cause. Finally, they should accept that, because subconscious reliance on stereotypes may be sensitive to the attributes and presentation of those being judged, individuals are not entirely powerless to affect the degree to which they will become victims of unconscious discrimination. On the other hand, those who would ignore the potential influence of group-based biases on human judgment should adopt a more agnostic stance.[277] They should acknowledge that there is a substantial body of experimental evidence that suggests that unconscious stereotyping can distort even the most seemingly fair and "neutral" judgments. They should accept that, because evidence of the workings of inadvertent bias is hard to generate and

276. Fiske, *supra* note 88, at 384.

277. The topic of unconscious discrimination rates little discussion and no index entries in recent books on race relations by conservative public intellectuals and academics. See DINESH D'SOUZA, THE END OF RACISM (1995); EPSTEIN, *supra* note 152; STEPHEN THERNSTROM & ABIGAIL THERNSTROM, AMERICA IN BLACK AND WHITE (1997).

unavoidably actuarial in form, abating its effects ought properly to be a collective responsibility. Finally, they should understand that creative, voluntary initiatives in the private sector, by perhaps making the law's heavy hand less necessary in the long run, may well promise the most social benefit at the least cost.